

Article

# Parallel Mechanism of Spectral Feature-Enhanced Maps in EEG-Based Cognitive Workload Classification

Yihong Zhang \* and Yuou Shen \*

College of Information Science and Technology, Engineering Research Center of Digitized Textile & Fashion Technology, Ministry of Education, DongHua University, Shanghai 201620, China

\* Correspondence: zhangyh@dhu.edu.cn (Y.Z.); 2171317@mail.dhu.edu.cn (Y.S.);

Tel.: +86-138-1782-6259 (Y.Z.); +86-150-0098-1298 (Y.S.)

Received: 25 December 2018; Accepted: 13 February 2019; Published: 16 February 2019



**Abstract:** Electroencephalography (EEG) provides a non-invasive, portable and low-cost way to convert neural signals into electrical signals. Using EEG to monitor people's cognitive workload means a lot, especially for tasks demanding high attention. Before deep neural networks became a research hotspot, the use of spectrum information and the common spatial pattern algorithm (CSP) was the most popular method to classify EEG-based cognitive workloads. Recently, spectral maps have been combined with deep neural networks to achieve a final accuracy of 91.1% across four levels of cognitive workload. In this study, a parallel mechanism of spectral feature-enhanced maps is proposed which enhances the expression of structural information that may be compressed by inter- and intra-subject differences. A public dataset and milestone neural networks, such as AlexNet, VGGNet, ResNet, DenseNet are used to measure the effectiveness of this approach. As a result, the classification accuracy is improved from 91.10% to 93.71%.

**Keywords:** cognitive workload; electroencephalography; spectral map; feature enhancement; neural network; mental workload

## 1. Introduction

Electroencephalography (EEG) data is composed of multi-channel bioelectric signals which are recorded by several electrodes and denote changes in bioelectric neural activity [1]. It is noninvasive and low cost and therefore used extensively in research. In the field of human psychology, EEG data is used to identify a persons' emotions, e.g., happiness, sadness, anger and shyness [2]. EEG data is also used in body control research to find changes in body movement (hands/feet-moving, blinking, etc.) and the corresponding brain activity [3–6]. This paper focuses on people's cognitive workload since physiological mental fatigue is a significant issue that impacts health and lifestyle [7,8]. In working environments, work stress can exceed the tolerance of people resulting in a decline in working performance. Having control of people's cognitive workload is important, particularly in activities which require a high level of concentration and attention. For example, engineers in a chemical factory must operate with high precision, and drivers need to maintain a healthy cognitive workload to prevent accidents [9]. Furthermore, workload estimations can be useful for many other applications, e.g., as quantitative measures for ergonomic and usability evaluations, or in medical applications [10].

Of the classical feature extraction methods, common space mode (CSP)-based methods are widely used to identify features [11–13], such as common sparse spectral pattern (CSSP) [14], sub-band common spatial pattern (SBCSP) [15], filter bank CSP (FBCSP) [16], discriminative CSP (DCSP) [17], and shrinkage regularized filter bank common spatial patterns (SR-FBCSP) [18] for binary EEG classification. Strong uncorrelated transform complex common spatial patterns (SUTCCSP) have

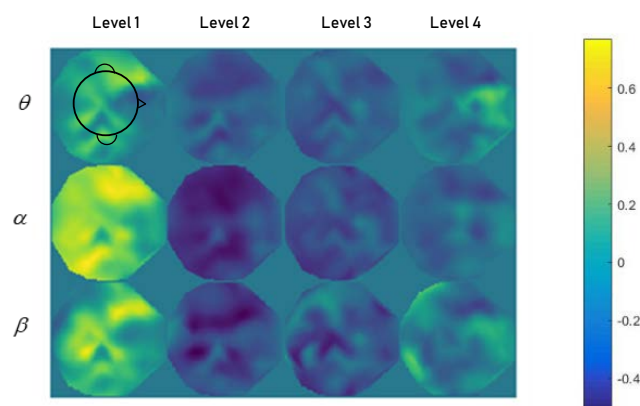
been used to identify different responses of the mu and beta rhythms of EEG traces corresponding to motor imagery tasks [19]. In addition to CSP-based methods, independent component analysis (ICA) is used to filter EEG data [20]. Pinheiro used C4.5 algorithms to control a virtual simulator using EEG signals [21]. In the workload classification area, Wang et al. used wireless EEG Signals to assess memory workload in the n-back task and achieved a general classification accuracy of over 80% [10]. Putze et al. trained and evaluated a Support Vector Machine (SVM)-based workload classifier and achieved a multimodal recognition rate of 95% for the two-class visual task which drops to 73% for cognitive tasks. For the full three-class problem, Putze et al. obtained recognition rates of 70% for the visual and 43% for the cognitive task [22]. Dominic et al. designed a SVM-based system for live recognition of mental workload and achieved an accuracy of 90.7 in the person-dependent evaluation while the accuracy dropped to 72.2 in the person independent evaluation [23]. Recently, Bashivan used Fast Fourier Transformation (FFT) to convert EEG data into the frequency domain and map the 3D spatial positions of electrodes to 2D, according to the distribution of the electrodes. The 4–30 Hz frequency domain containing prime cortical activity is divided into theta (4–7Hz), alpha (8–13Hz) and beta (13–30Hz) frequency bands, generating 3-channel spectral maps which are sent to deep neural networks. This method combines EEG data with deep neural networks and achieve an accuracy of 91.1% [24]. However, the data of the test and training set originated from one cohort of 13 subjects, and therefore if the model was to be used to classify their cognitive workload level of a new volunteer, the accuracy could decrease. Furthermore, FFT was performed on the time series for each trial to estimate the power spectrum of the signal. The sum of squared absolute values within  $\theta$ ,  $\alpha$  and  $\beta$  frequency bands was computed and used as separate measurements for each electrode, which causes different spectral ranges of spectral maps from different subjects different trails (the times subject finishing experimental tasks calls trails) [24]. This method extracts information from the three frequency bands and this significantly reduces the quantity of data, however it also compresses the structure information of each trail.

This study proposes a novel approach to enhance the features of spectral maps and apply some milestone achievements in deep learning to the EEG-based workload classification area. In the feature enhancement part, 3-channel feature-enhanced maps are obtained and combined with original 3-channel spectral maps, thus building the parallel mechanism of spectral feature-enhanced maps which have 6-channels. In the feature extraction part, the shortcut connection in ResNet and short paths in DenseNet improve the ability of feature extraction from the spectral map, and are undertaken for the first time in the EEG-based cognitive workload classification area [25,26]. The final accuracy of cognitive workload classification using a public dataset is improved from 91.1% to 93.6% and the generalization of the network is also improved.

## 2. EEG Dataset and Physiological Underpinnings

In order to compare the effects of the parallel mechanism of the spectral feature-enhanced (PMSFE) maps with spectral maps, the public database provided by Bashivan on GitHub was used in our experiments (Web: <https://github.com/pbashivan/EEGLearn>). The dataset records 2670 trails of cognitive workload records from 13 subjects whose age ranged from 24 to 33, obtained with 64 electrodes placed over the scalp at standard 10-10 locations. The method of testing subjects' cognitive workload was similar to Sternberg's memory experiment [27,28]. At the beginning, the participants see a number of English letters that are circularly distributed at the same angle in the center of the screen. The letters on the screen are continuously displayed for 0.5 s and participants need to remember these words in 0.5 s. After the letter disappears, participants need to judge whether a random letter appeared after three seconds' margin. If participants judge correctly, it will be recorded as a valid trail. A total of 2670 trails are recorded in the database and divided into four load levels according to the number of memory letters 2, 4, 6 and 8, respectively. The higher the number of memories, the higher the corresponding load level.

Physiological studies show that prefrontal cortex is the mostly involved during cognitive and memory tasks [29], which is consistent with the features shown in Figure 1, which illustrates the mean topological spectral maps calculated for the  $\theta$ ,  $\alpha$  and  $\beta$  frequency bands of different load levels. The color in that prefrontal cortex in level 1 bias blue in all the  $\theta$ ,  $\alpha$  and  $\beta$  frequency bands. As the difficulty of memory tasks increases, this area is getting brighter and brighter. What's more, higher load levels enhance the power in higher frequency bands [30]. The spectral maps at level 2, level 3 and level 4 generally conform to physiological theory except level 1. This is because participants can manage simple memory tasks without paying too much attention. They may didn't plunge themselves in this task and many ideas which can be considered as noise in daily life may by going on in their mind. At level 2 and higher levels, they have to focus on the task or when they can't manage the task, then they reduce unnecessary thoughts and the spectral map is helped according to physiological theory.



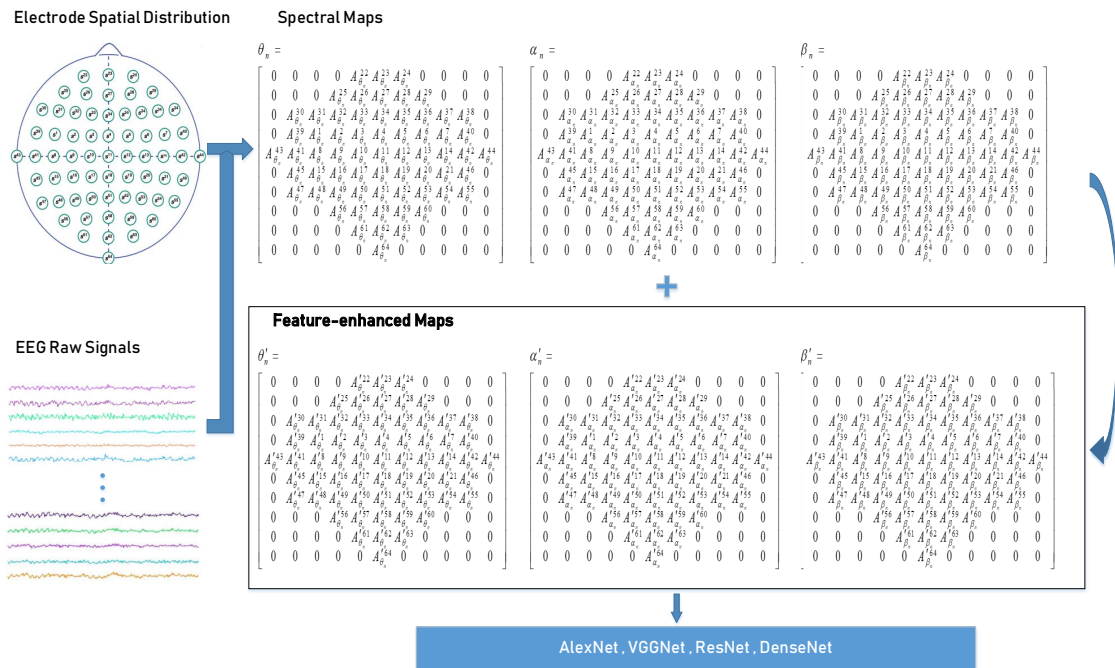
**Figure 1.** The mean topological spectral maps of  $\theta$ ,  $\alpha$  and  $\beta$  at a 4 classes load level. Maps in the first line show the spectral changes in the  $\theta$  frequency band under different load level. Maps in the second line show the spectral changes in the  $\alpha$  frequency band under different load levels. Maps in the third line show the spectral changes in the  $\beta$  frequency band under different load levels. All values have been subtracted from the average spectral value of the whole dataset.

### 3. Parallel Mechanism of Spectral Feature-Enhanced Maps

#### 3.1. The Structure of Parallel Mechanism

Figure 2 shows the overview of our approach. EEG raw data from different channels is converted into the frequency domain using FFT and combined with the electrode space information, producing  $\theta$ ,  $\alpha$  and  $\beta$  spectral maps [24]. After performing the feature enhancement process, the overall structure features are determined as  $\theta'$ ,  $\alpha'$  and  $\beta'$  spectral-enhanced maps.

$\theta$ ,  $\theta'$ ,  $\alpha$ ,  $\alpha'$ ,  $\beta$  and  $\beta'$  form 6-channel spectral maps, which is the establishment of a parallel mechanism. Finally, PMSFE maps are sent to several milestone neural networks, e.g., AlexNet, VGGNet, ResNet, DenseNet to test for accuracy and generalization [25,26,31–34].



**Figure 2.** An overview of the approach. ( $\theta_n$ ,  $\alpha_n$  and  $\beta_n$  represent the spectral maps at 4–7 Hz, 8–13 Hz, and 13–30 Hz of the No.  $n$  trail;  $A^1_{\theta_n} \dots A^{64}_{\theta_n}$ ,  $A^1_{\alpha_n} \dots A^{64}_{\alpha_n}$ ,  $A^1_{\beta_n} \dots A^{64}_{\beta_n}$  represent the spectral values recorded by the No. 1 channel electrode to No. 64 channel electrode at 4–7 Hz, 8–13 Hz, and 13–30 Hz of the No.  $n$  trail).

### 3.2. The Development of PMSFE Maps

The first step is to get the spectral values and form corresponding vectors [24,29]. The 1D vectors of spectral values from the  $\theta$ ,  $\alpha$  and  $\beta$  frequency bands are shown as follows:

$$\theta_n = [A^1_{\theta_n}, A^2_{\theta_n}, A^3_{\theta_n}, A^i_{\theta_n} \dots, A^{64}_{\theta_n}]^T \tag{1}$$

$$\alpha_n = [A^1_{\alpha_n}, A^2_{\alpha_n}, A^3_{\alpha_n}, A^i_{\alpha_n} \dots, A^{64}_{\alpha_n}]^T \tag{2}$$

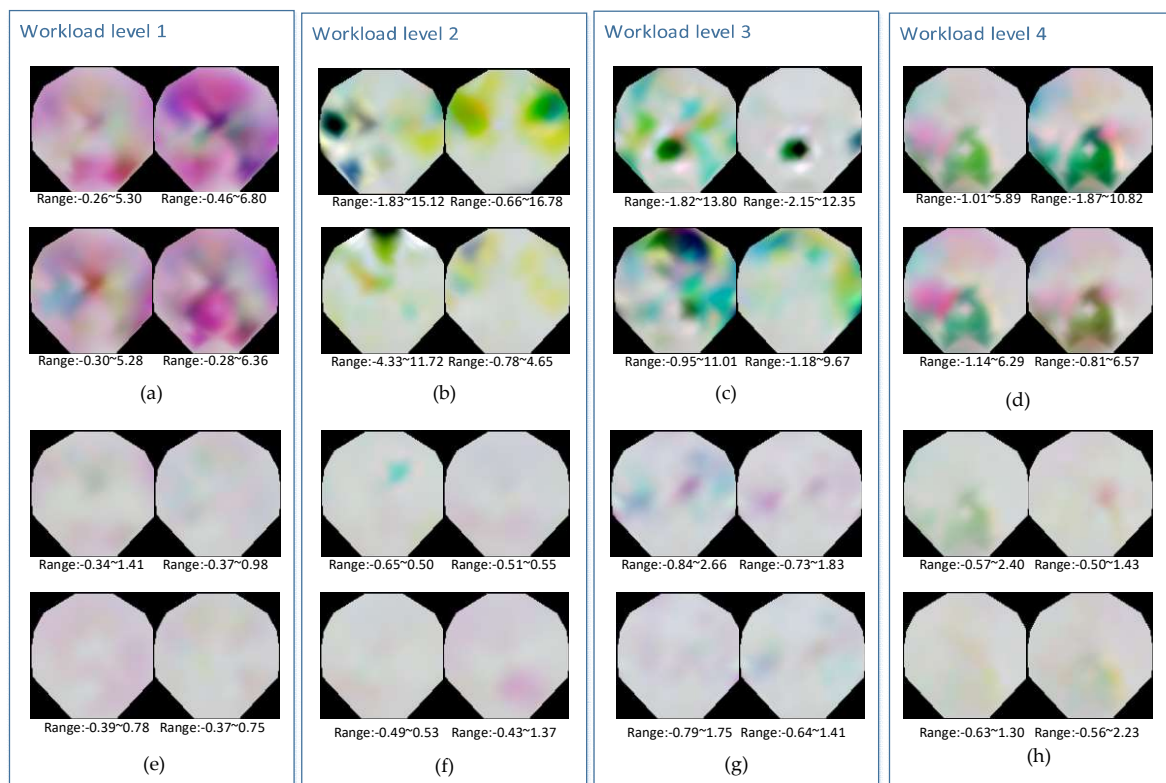
$$\beta_n = [A^1_{\beta_n}, A^2_{\beta_n}, A^3_{\beta_n}, A^i_{\beta_n} \dots, A^{64}_{\beta_n}]^T \tag{3}$$

$A^i_{\theta_n}$ ,  $A^i_{\alpha_n}$ ,  $A^i_{\beta_n}$  represent the spectral values recorded by the No.  $i$  channel electrode at 4–7 Hz, 8–13 Hz and 13–30 Hz of the No.  $n$  trail. After considering the electrode spatial distribution, 2D vectors are produced:

$$\theta_n / \alpha_n / \beta_n = \begin{bmatrix} 0 & 0 & 0 & A^{22}_{\theta_n / \alpha_n / \beta_n} & A^{23}_{\theta_n / \alpha_n / \beta_n} & A^{24}_{\theta_n / \alpha_n / \beta_n} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & A^{25}_{\theta_n / \alpha_n / \beta_n} & A^{26}_{\theta_n / \alpha_n / \beta_n} & A^{27}_{\theta_n / \alpha_n / \beta_n} & A^{28}_{\theta_n / \alpha_n / \beta_n} & A^{29}_{\theta_n / \alpha_n / \beta_n} & 0 & 0 & 0 \\ 0 & A^{30}_{\theta_n / \alpha_n / \beta_n} & A^{31}_{\theta_n / \alpha_n / \beta_n} & A^{32}_{\theta_n / \alpha_n / \beta_n} & A^{33}_{\theta_n / \alpha_n / \beta_n} & A^{34}_{\theta_n / \alpha_n / \beta_n} & A^{35}_{\theta_n / \alpha_n / \beta_n} & A^{36}_{\theta_n / \alpha_n / \beta_n} & A^{37}_{\theta_n / \alpha_n / \beta_n} & A^{38}_{\theta_n / \alpha_n / \beta_n} & 0 \\ 0 & A^{39}_{\theta_n / \alpha_n / \beta_n} & A^1_{\theta_n / \alpha_n / \beta_n} & A^2_{\theta_n / \alpha_n / \beta_n} & A^3_{\theta_n / \alpha_n / \beta_n} & A^4_{\theta_n / \alpha_n / \beta_n} & A^5_{\theta_n / \alpha_n / \beta_n} & A^6_{\theta_n / \alpha_n / \beta_n} & A^7_{\theta_n / \alpha_n / \beta_n} & A^{40}_{\theta_n / \alpha_n / \beta_n} & 0 \\ A^{43}_{\theta_n / \alpha_n / \beta_n} & A^{41}_{\theta_n / \alpha_n / \beta_n} & A^8_{\theta_n / \alpha_n / \beta_n} & A^9_{\theta_n / \alpha_n / \beta_n} & A^{10}_{\theta_n / \alpha_n / \beta_n} & A^{11}_{\theta_n / \alpha_n / \beta_n} & A^{12}_{\theta_n / \alpha_n / \beta_n} & A^{13}_{\theta_n / \alpha_n / \beta_n} & A^{14}_{\theta_n / \alpha_n / \beta_n} & A^{42}_{\theta_n / \alpha_n / \beta_n} & A^{44}_{\theta_n / \alpha_n / \beta_n} \\ 0 & A^{45}_{\theta_n / \alpha_n / \beta_n} & A^{15}_{\theta_n / \alpha_n / \beta_n} & A^{16}_{\theta_n / \alpha_n / \beta_n} & A^{17}_{\theta_n / \alpha_n / \beta_n} & A^{18}_{\theta_n / \alpha_n / \beta_n} & A^{19}_{\theta_n / \alpha_n / \beta_n} & A^{20}_{\theta_n / \alpha_n / \beta_n} & A^{21}_{\theta_n / \alpha_n / \beta_n} & A^{46}_{\theta_n / \alpha_n / \beta_n} & 0 \\ 0 & A^{47}_{\theta_n / \alpha_n / \beta_n} & A^{48}_{\theta_n / \alpha_n / \beta_n} & A^0_{\theta_n / \alpha_n / \beta_n} & A^{49}_{\theta_n / \alpha_n / \beta_n} & A^{50}_{\theta_n / \alpha_n / \beta_n} & A^{51}_{\theta_n / \alpha_n / \beta_n} & A^{52}_{\theta_n / \alpha_n / \beta_n} & A^{53}_{\theta_n / \alpha_n / \beta_n} & A^{54}_{\theta_n / \alpha_n / \beta_n} & 0 \\ 0 & 0 & 0 & A^{56}_{\theta_n / \alpha_n / \beta_n} & A^{57}_{\theta_n / \alpha_n / \beta_n} & A^{58}_{\theta_n / \alpha_n / \beta_n} & A^{59}_{\theta_n / \alpha_n / \beta_n} & A^{60}_{\theta_n / \alpha_n / \beta_n} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & A^{59}_{\theta_n / \alpha_n / \beta_n} & A^{61}_{\theta_n / \alpha_n / \beta_n} & A^{62}_{\theta_n / \alpha_n / \beta_n} & A^{63}_{\theta_n / \alpha_n / \beta_n} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & A^{61}_{\theta_n / \alpha_n / \beta_n} & A^{62}_{\theta_n / \alpha_n / \beta_n} & A^{63}_{\theta_n / \alpha_n / \beta_n} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & A^{64}_{\theta_n / \alpha_n / \beta_n} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{4}$$

These 3-channel maps are delivered into RGB channels and the visualization pictures are shown in Figure 3. Significant features are identified in the upper parts of the four workload levels, while features in the lower parts are not obvious. This could be due to the colorful pictures which indicate that trails have a wider range of maximum and minimum values, while pictures with unremarkable features

don't have a large fluctuation range of spectral power values. In the four cognitive load levels, pictures with unremarkable features are very common, indicating the 2670 trails range should be considered and every trail needs to be enhanced to fully identify the structural information. If these spectral maps are only regulated in the whole standard, the structure information of maps with unremarkable features would be compressed. Since EEG data is very sensitive, any slight body movement or distraction of mind could result in significant changes in the spectral map and influence the final result. Therefore, in addition to the power information of each electrode frequency domain, the whole structure of the spectral map also needs to be considered. An individual's cognitive workload level is related to its spectral power and also the overall spatial distribution. Spectral maps that have not been enhanced on a self-based basis can weaken the expression of the structure information in deep neural networks.



**Figure 3.** Visualization of examples from four different workload levels. (a–d) show some samples of maps with obvious features in four classes and the range of spectral value are given. (e–h) show some samples of maps with unobvious features in four classes and the range of spectral value are given. All these values have been subtracted from the average spectral value of the whole dataset.

The details of enhancing the structure features of spectral map are described as follows. After the  $\theta_n$ ,  $\alpha_n$  and  $\beta_n$  spectral power values have been subtracted from the overall average, Equation (5) is used to project all the  $\theta_n$ ,  $\alpha_n$  and  $\beta_n$  into a range from 0 to 1, and  $\theta_n^*$ ,  $\alpha_n^*$  and  $\beta_n^*$  are thus obtained:

$$A_n^{i*} = \frac{1}{1 + e^{-0.15 \cdot A_n^i}} \quad (5)$$

Since the spectral power is calculated by the sum of squared absolute values within each of the  $\theta_n$ ,  $\alpha_n$  and  $\beta_n$  frequency bands, Equation (5) could weaken the impact of the square operation and avoid the interference from abnormal large values. Equation (8) is used to widen the range of minimum and maximum values:

$$\text{Max}(A_n^{i*}) = \text{Max}(A_{\theta_n}^{1*}, A_{\theta_n}^{2*} \dots, A_{\theta_n}^{i*} \dots, A_{\theta_n}^{64*}, A_{\alpha_n}^{1*}, A_{\alpha_n}^{2*} \dots, A_{\alpha_n}^{i*} \dots, A_{\alpha_n}^{64*}, A_{\beta_n}^{1*}, A_{\beta_n}^{2*} \dots, A_{\beta_n}^{i*} \dots, A_{\beta_n}^{64*}) \quad (6)$$

$$\text{Min}(A_n^{i*}) = \text{Min}(A_{\theta_n}^{1*}, A_{\theta_n}^{2*} \dots, A_{\theta_n}^{i*} \dots, A_{\theta_n}^{64*}, A_{\alpha_n}^{1*}, A_{\alpha_n}^{2*} \dots, A_{\alpha_n}^{i*} \dots, A_{\alpha_n}^{64*}, A_{\beta_n}^{1*}, A_{\beta_n}^{2*} \dots, A_{\beta_n}^{i*} \dots, A_{\beta_n}^{64*}) \quad (7)$$

$$A_n^i = \frac{A_n^{i*} - \text{Min}(A_n^{i*})}{\text{Max}(A_n^{i*}) - \text{Min}(A_n^{i*})} \quad (8)$$

$\text{Max}(A_n^{i*})$  is the maximum value of  $\theta_n^*$ ,  $\alpha_n^*$  and  $\beta_n^*$  in the No. n trail;  $\text{Min}(A_n^{i*})$  is minimum value of  $\theta_n^*$ ,  $\alpha_n^*$  and  $\beta_n^*$  the in No. n trail.

Substituting Equation (5) into Equation (8), the final feature-enhanced equation and spectral feature-enhanced maps are obtained as follows:

$$A_n^i = \frac{(e^{\text{Min}(-0.15A_n^i)} - e^{-0.15A_n^i}) \cdot (1 + e^{\text{Max}(-0.15A_n^i)})}{(e^{\text{Min}(-0.15A_n^i)} - e^{\text{Max}(-0.15A_n^i)}) \cdot (1 + e^{-0.15A_n^i})} \quad (9)$$

$$\theta_n' / \alpha_n' / \beta_n' = \begin{bmatrix} 0 & 0 & 0 & 0 & A_{\theta_n/\alpha_n/\beta_n}^{22} & A_{\theta_n/\alpha_n/\beta_n}^{23} & A_{\theta_n/\alpha_n/\beta_n}^{24} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & A_{\theta_n/\alpha_n/\beta_n}^{25} & A_{\theta_n/\alpha_n/\beta_n}^{26} & A_{\theta_n/\alpha_n/\beta_n}^{27} & A_{\theta_n/\alpha_n/\beta_n}^{28} & A_{\theta_n/\alpha_n/\beta_n}^{29} & 0 & 0 & 0 \\ 0 & A_{\theta_n/\alpha_n/\beta_n}^{30} & A_{\theta_n/\alpha_n/\beta_n}^{31} & A_{\theta_n/\alpha_n/\beta_n}^{32} & A_{\theta_n/\alpha_n/\beta_n}^{33} & A_{\theta_n/\alpha_n/\beta_n}^{34} & A_{\theta_n/\alpha_n/\beta_n}^{35} & A_{\theta_n/\alpha_n/\beta_n}^{36} & A_{\theta_n/\alpha_n/\beta_n}^{37} & A_{\theta_n/\alpha_n/\beta_n}^{38} & 0 \\ 0 & A_{\theta_n/\alpha_n/\beta_n}^{39} & A_{\theta_n/\alpha_n/\beta_n}^{40} & A_{\theta_n/\alpha_n/\beta_n}^{41} & A_{\theta_n/\alpha_n/\beta_n}^{42} & A_{\theta_n/\alpha_n/\beta_n}^{43} & A_{\theta_n/\alpha_n/\beta_n}^{44} & A_{\theta_n/\alpha_n/\beta_n}^{45} & A_{\theta_n/\alpha_n/\beta_n}^{46} & A_{\theta_n/\alpha_n/\beta_n}^{47} & 0 \\ A_{\theta_n/\alpha_n/\beta_n}^{43} & A_{\theta_n/\alpha_n/\beta_n}^{41} & A_{\theta_n/\alpha_n/\beta_n}^{38} & A_{\theta_n/\alpha_n/\beta_n}^{9} & A_{\theta_n/\alpha_n/\beta_n}^{10} & A_{\theta_n/\alpha_n/\beta_n}^{11} & A_{\theta_n/\alpha_n/\beta_n}^{12} & A_{\theta_n/\alpha_n/\beta_n}^{13} & A_{\theta_n/\alpha_n/\beta_n}^{14} & A_{\theta_n/\alpha_n/\beta_n}^{15} & A_{\theta_n/\alpha_n/\beta_n}^{16} \\ 0 & A_{\theta_n/\alpha_n/\beta_n}^{45} & A_{\theta_n/\alpha_n/\beta_n}^{15} & A_{\theta_n/\alpha_n/\beta_n}^{16} & A_{\theta_n/\alpha_n/\beta_n}^{17} & A_{\theta_n/\alpha_n/\beta_n}^{18} & A_{\theta_n/\alpha_n/\beta_n}^{19} & A_{\theta_n/\alpha_n/\beta_n}^{20} & A_{\theta_n/\alpha_n/\beta_n}^{21} & A_{\theta_n/\alpha_n/\beta_n}^{46} & 0 \\ 0 & A_{\theta_n/\alpha_n/\beta_n}^{47} & A_{\theta_n/\alpha_n/\beta_n}^{18} & A_{\theta_n/\alpha_n/\beta_n}^{19} & A_{\theta_n/\alpha_n/\beta_n}^{50} & A_{\theta_n/\alpha_n/\beta_n}^{51} & A_{\theta_n/\alpha_n/\beta_n}^{52} & A_{\theta_n/\alpha_n/\beta_n}^{53} & A_{\theta_n/\alpha_n/\beta_n}^{54} & A_{\theta_n/\alpha_n/\beta_n}^{55} & 0 \\ 0 & 0 & 0 & A_{\theta_n/\alpha_n/\beta_n}^{56} & A_{\theta_n/\alpha_n/\beta_n}^{57} & A_{\theta_n/\alpha_n/\beta_n}^{58} & A_{\theta_n/\alpha_n/\beta_n}^{59} & A_{\theta_n/\alpha_n/\beta_n}^{60} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & A_{\theta_n/\alpha_n/\beta_n}^{61} & A_{\theta_n/\alpha_n/\beta_n}^{62} & A_{\theta_n/\alpha_n/\beta_n}^{63} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & A_{\theta_n/\alpha_n/\beta_n}^{64} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (10)$$

The original 3-channel spectral maps  $\theta_n$ ,  $\alpha_n$  and  $\beta_n$  and the feature-enhanced 3-channel maps  $\theta_n'$ ,  $\alpha_n'$  and  $\beta_n'$  form the 6-channel maps, which build the PMSFE maps since the sum of squared absolute values within each of the three frequency bands was computed and used as separate measurements for each electrode [24]. This way of calculation can extract main information effectively and realize real-time processing but may cause different ranges of spectral value. Equation (9) can reduce the interference from extreme values and improve the expression of compressed structural information. Figure 4 contrasts the 3D images of the  $\theta$ ,  $\alpha$  and  $\beta$  channels and the enhanced  $\theta'$ ,  $\alpha'$  and  $\beta'$  channels and the explanation of mentioned notations are given in Table 1.

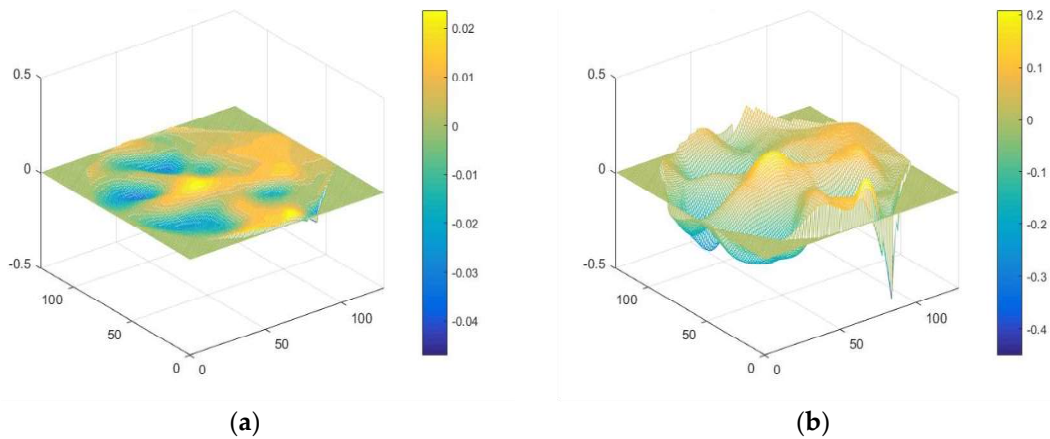
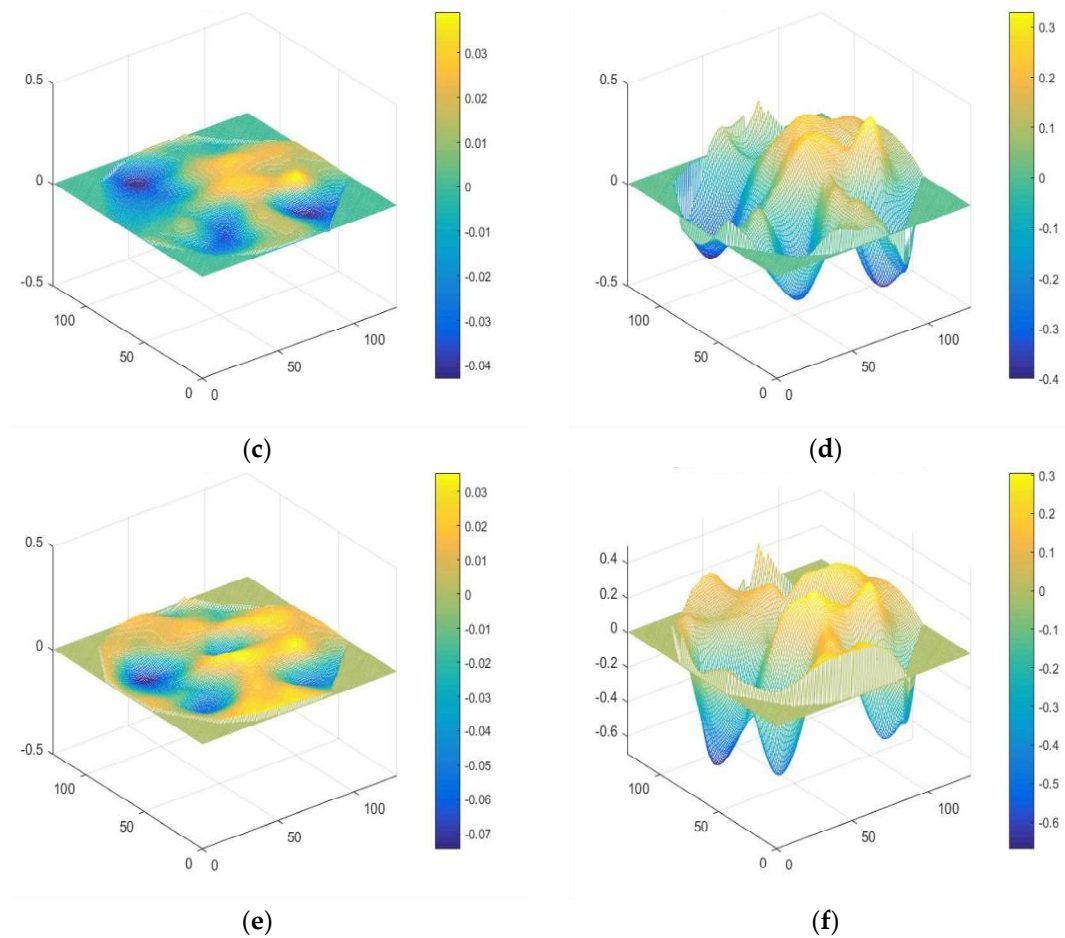


Figure 4. Cont.



**Figure 4.** (a) 3D image of the  $\theta$  channel; (b) 3D image of the feature-enhanced  $\theta'$  channel; (c) 3D image of the  $\alpha$  channel; (d) 3D image of the feature-enhanced  $\alpha'$  channel; (e) 3D image of the  $\beta$  channel; (f) 3D image of the feature-enhanced  $\beta'$  channel.

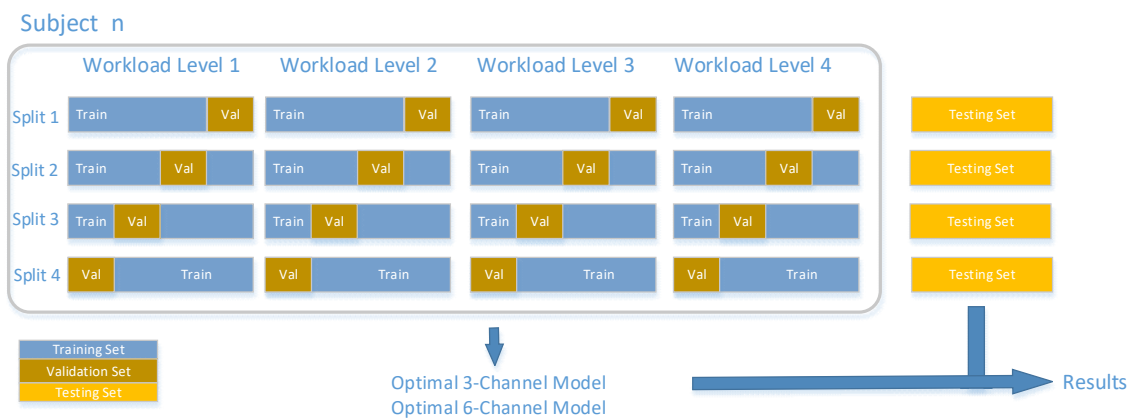
**Table 1.** The explanation of variables.

Notation	Explanation
$\theta_n$	The spectral values in the frequency bands of theta (4–7 Hz)
$\alpha_n$	The spectral values in the frequency bands of alpha (8–13 Hz)
$\beta_n$	The spectral values in the frequency bands of beta (13–30 Hz)
$A_n^i / A_{\theta_n, \alpha_n, \beta_n}^i$	The spectral values recorded by No. $i$ channel electrode at 4–7 Hz, 8–13 Hz and 13–30 Hz of the No. $n$ trail
$A_n^{i*} / A_{\theta_n, \alpha_n, \beta_n}^{i*}$	The spectral value in No. $i$ channel electrode at 4–7 Hz, 8–13 Hz and 13–30 Hz of the No. $n$ trail processed by Equation (5)
$A_n^i / A_{\theta_n, \alpha_n, \beta_n}^i$	The final enhanced spectral value in No. $i$ channel electrode at 4–7 Hz, 8–13 Hz and 13–30 Hz of the No. $n$ trail processed by Equation (9)
$\theta_n^*$	The spectral values in the $\theta$ (4–7 Hz) frequency bands processed by Equation (5)
$\alpha_n^*$	The spectral values in the $\alpha$ (8–13 Hz) frequency bands processed by Equation (5)
$\beta_n^*$	The spectral values in the $\beta$ (13–30 Hz) frequency bands processed by Equation (5)
$Max(A_n^{i*})$	The maximum value of $\theta_n^*$ , $\alpha_n^*$ and $\beta_n^*$ in the No. $n$ trail
$Min(A_n^{i*})$	The minimum value of $\theta_n^*$ , $\alpha_n^*$ and $\beta_n^*$ in the No. $n$ trail
$\theta'_n$	The enhanced $\theta$ (4–7 Hz) frequency bands
$\alpha'_n$	The enhanced $\alpha$ (8–13 Hz) frequency bands
$\beta'_n$	The enhanced $\beta$ (13–30 Hz) frequency bands

#### 4. Baseline Methods and Training Process

To measure the effect of PMSFE maps compared to unenhanced spectral maps, most popular and milestone deep neural networks, such as AlexNet, VGGNet, ResNet, and DenseNet are used as baseline methods [25,26,31–34]. In the feature-extracting layers,  $\theta$ ,  $\alpha$ ,  $\beta$  and  $\theta'$ ,  $\alpha'$ ,  $\beta'$  are trained separately and do not affect each other. In the classification layer, they are fully linked with each other.

The proportion of training set, validation set and test set is 6:2:2 which is randomly divided according to every workload level of every subject proportionately. Stratified cross validation is used to find the optimal model [35]. Part (80%) of the whole data set is divided into four folds. We take turns to use each fold as the validation set and use the remaining three folds as the training set (Figure 5). Every split can get the error of the model and the average error of the four errors indicates the model performance. By adjusting the hyperparameters, an optimal 3-channel model and 6-channel model can be found. The matrix of the optimal model is saved and transferred when using the testing set. The training set is trained based on Pytorch on an Nvidia GTX-1080TI GPU. During the process of training, transfer learning with a learning rate of 0.001 and momentum 0.9 is used to reduce the training time [36–38].



**Figure 5.** The process of stratified 4-fold cross validation and final testing. Training set and validation set accounts for 80% which is used to find the optimal 3-channel model and 6-channel model. Then the remaining 20% testing set was used to calculate the final accuracy.

Some basic information and parameter settings are introduced and the corresponding validation loss line during training are given correspondingly. The convergence speed and state of neural networks can be directly observed through the change of validation loss.

##### 4.1. AlexNet

Due to limited GPU computing capacity in its day, AlexNet was divided into two parts. Today, improved GPU computing power has solved the problem and some slight changes have been made. In the feature learning section, Bashivan's method is used to form the topology-preserving spectral maps in  $224 \times 224$  size [24]. The three input channels are extended into six channels, the size of kernel in five convolutional layers is  $11 \times 11$ ,  $5 \times 5$ ,  $3 \times 3$ ,  $3 \times 3$  and  $3 \times 3$  which means the sizes of the matrices used to convolve with maps and the corresponding numbers are 64, 192, 384, 256, 256. Usually more kernels exist in latter layers to ensure features of formal layer being fully abstracted. In the classification section, the input and output parameters of the last linear are changed to 4096 and 4 (4096 are the output number of formal layer and 4 is determined by the number of classes). Figure 6 shows the validation loss of spectral maps and PMSFE maps over training set using AlexNet.

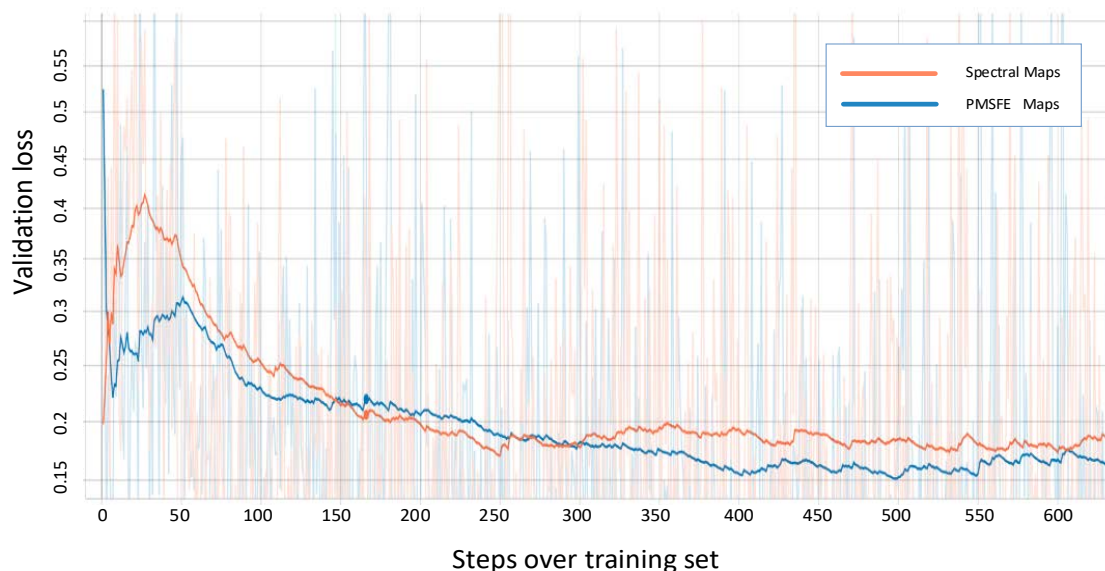




**Figure 6.** Validation loss of spectral maps and PMSFE maps over training set using AlexNet. (Smooth = 0.95 in Tensorboardx).

#### 4.2. VGGNet

On the basis of AlexNet, VGGNet deepened the neural network and the convolutional layers using the same size convolutional filter  $3 \times 3$ , achieving greater nonlinearity with less parameters. The vision of VGGNet displayed in the Figure 7 is VGGNet-19. In the feature learning section, the number of kernels in 13 convolutional layers is 64, 64, 128, 128, 256, 256, 256, 512, 512, 512, 512, 512 and 512. The change of classification section is the same in AlexNet.

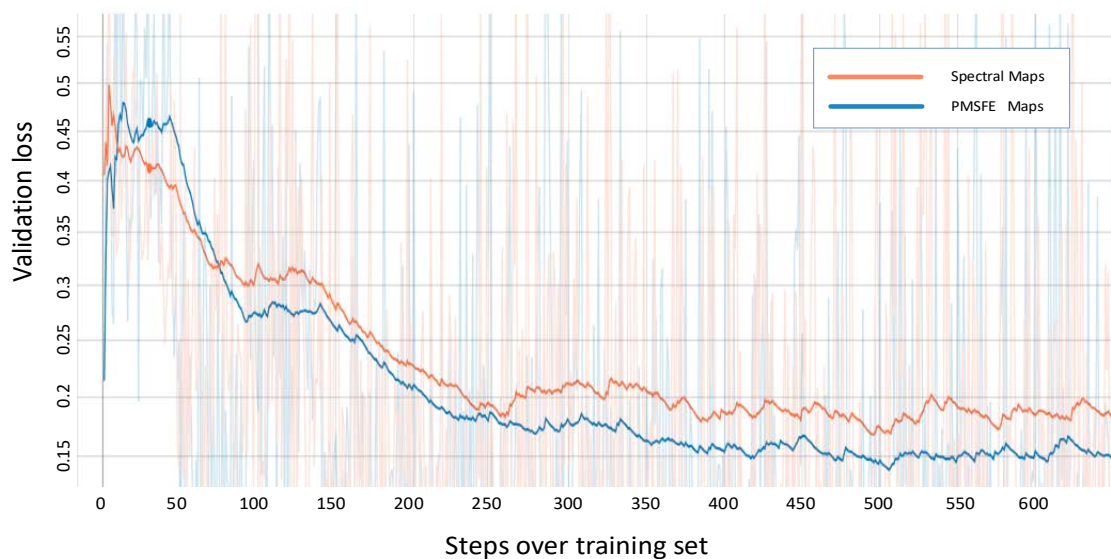


**Figure 7.** Validation loss of spectral maps and PMSFE maps over the training set in VGGNet-19 (Smooth = 0.97 in Tensorboardx).

#### 4.3. ResNet

Before the ResNet unit appeared, gradient disappearance and gradient explosions could limit traditional deep networks from being well-trained. In its structure, each block of ResNet consisted of a series of layers and a shortcut connection which connect the input and output of the module

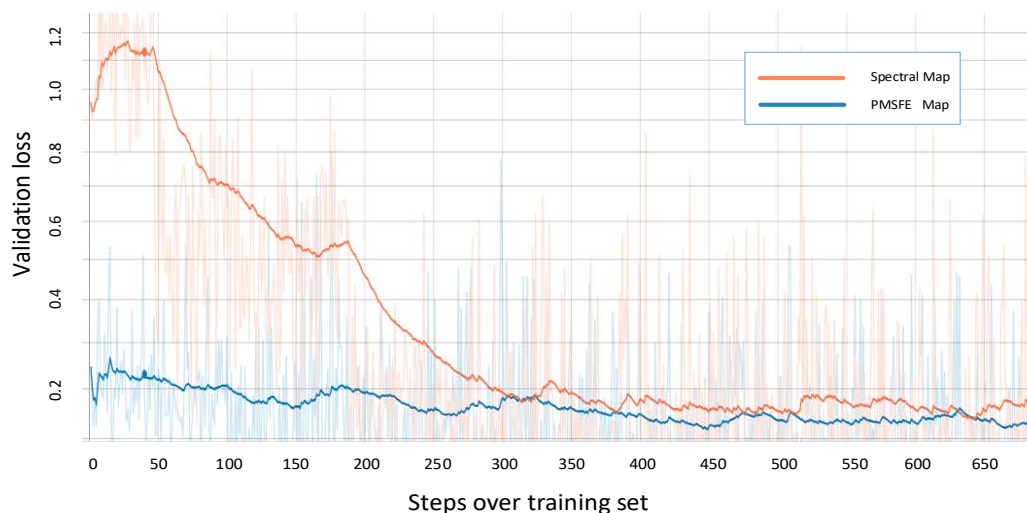
and perform the addition operation, addressing gradient disappearance. The validation loss in ResnetV2-152 is recorded as an example to demonstrate the difference between spectral maps and PMSFE maps in Figure 8 and the output of its fully connected layers is also changed to 4.



**Figure 8.** Validation loss of spectral maps and PMSFE maps over training set in ResNet-152. (Smooth = 0.987 in Tensorboardx).

#### 4.4. DenseNet

Unlike ResNet, which adds the output to the input to form a residual structure, DenseNet parallels the output to the input, enabling each layer to directly get the output of all the previous layers. Each layer of DenseNet only requires a few learnt features, significantly reducing the parameters and computation required. The validation loss in Densenet-201 is recorded as an example to demonstrate the difference between spectral maps and PMSFE maps in Figure 9 and the output of its fully connected layers is also changed to 4.



**Figure 9.** Validation loss of spectral maps and PMSFE maps over training set in DenseNet-201. (Smooth = 0.978 in Tensorboardx).

The two colored lines represent the validation loss of PMSFE maps and spectral maps in the four kinds of network and the cross-entropy loss function is used to calculate the loss value.

All four validation loss lines demonstrate that PMSFE maps have a lower validation loss compared to spectral maps and the difference ranges from 0.010 to 0.022, indicating that the parallel mechanism creates a smoother training process and eventually stabilizes at a lower level.

## 5. Results and Discussions

The effect of a parallel mechanism is tested from two aspects. In the first aspect, the accuracy of Spectral Maps and PMSFE maps are calculated according to four trained models: AlexNet, VGGNet, ResNet and DenseNet. The second aspect is the testing of generalizations between spectral and PMSFE maps in neural networks.

### 5.1. Accuracy

Table 2 shows the final accuracy of spectral maps and PMSFE maps in nine models, and the increased rates after using a parallel mechanism. By adding feature-enhanced maps into the network the accuracy of AlexNet and VGGNet are increased from 88.77% to 90.18%, from 91.46% to 92.77% in 16-layers vision with batch normalization (BN) and from 91.70% to 93.34% in 19-layers vision with BN [39]. Although the mean accuracy of VGGNet performs better than AlexNet, the rate of increase is similar (approximately 1.5%).

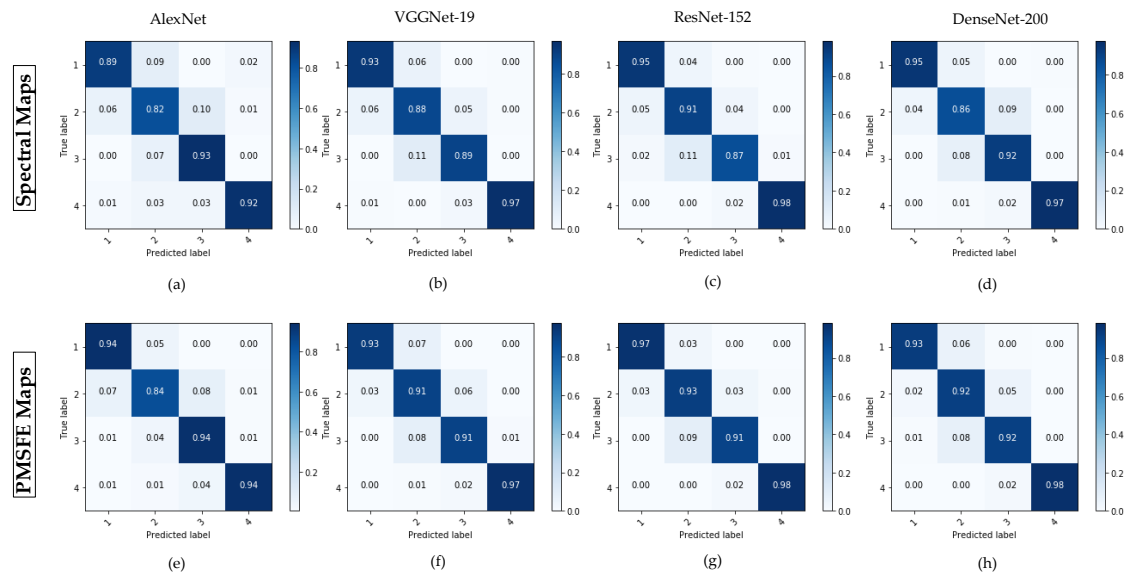
The rate of increase is approximately 1% in ResNet and 0.56% in DenseNet, due to the feature-enhanced map enhancing the structure information using every trail. This phenomenon also reflects the ability of DenseNet and ResNet to extract features better than AlexNet's and VGGNet's.

The highest accuracy is 93.71% using PMSFE maps in ResNet-152, which increases the final accuracy from 91.1% to 93.71%. Across the rate increase of 2.61%, 1.52% is due to performance improvements of the network, and 1.09% is due to the parallel mechanism of spectral feature-enhanced maps.

**Table 2.** The accuracy of spectral maps and PMSFE maps across all networks and the rate of increase after using PMSFE maps.

Network	Spectral Maps		PMSFE Maps		Increased Rate (%)
	Mean Accuracy (%)	Standard Deviation	Mean Accuracy (%)	Standard Deviation	
AlexNet	88.78	0.0019	90.18	0.0031	1.4
VGGNet-16BN	91.46	0.0026	92.77	0.0030	1.31
VGGNet-19BN	91.70	0.0016	93.34	0.0026	1.64
ResNet-50	92.50	0.0027	93.08	0.0023	0.58
ResNet-101	92.61	0.0025	93.68	0.0032	1.07
ResNet-152	92.62	0.0023	93.71	0.0027	1.09
DenseNet-121	92.33	0.0051	92.87	0.0039	0.54
DenseNet-169	92.56	0.0029	93.08	0.0027	0.52
DenseNet-201	92.88	0.0028	93.47	0.0021	0.59

Figure 10 shows the confusion matrix of spectral maps and PMSFE maps in AlexNet, VGGNet-19, ResNet-152 and DenseNet-200. It obvious that load level 2 and load level 3 are easily confused when using original spectral maps. The probabilities of the neural network misclassifying level 3 into level 2 are correspondingly 7%, 11%, 11% and 8%. The percentages decrease to 4%, 8%, 9% and 8% after using PMSFE maps.



**Figure 10.** The confusion matrix of spectral maps and PMSFE maps. (a) The confusion matrix of spectral maps in AlexNet. (b) The confusion matrix of spectral maps in VGGNet-19. (c) The confusion matrix of spectral maps in ResNet-152. (d) The confusion matrix of spectral maps in DenseNet-200. (e) The confusion matrix of PMSFE maps in AlexNet. (f) The confusion matrix of PMSFE maps in VGGNet-19. (g) The confusion matrix of PMSFE maps in ResNet-152. (h) The confusion matrix of PMSFE maps in DenseNet-200.

## 5.2. Generalization

In actual usage of cognitive workload classification, the final accuracy of new subjects is not necessarily so high. Although the training and test set don't overlap, they are all from the same 13 subjects, which results in features of specific subjects being already trained by deep neural networks and the accuracy would be higher. In the testing of generalization after using PMSFE maps, trails of specific subjects are identified as the test set, and these trails from the remaining 12 subjects is used as training set. For example, in Table 3 the EEG data from subject No.1 is not included when training the network and is only used as the test set. This method simulates the actual application and guarantees the validity generalization.

**Table 3.** The accuracy of spectral maps and PMSFE maps tested on 13 subjects and the rate of increase after using PMSFE maps.

Subject	Spectral Maps (%)	PMSFE Maps (%)	Increased Rate (%)
1	55.68	58.38	2.70
2	70.28	72.17	1.89
3	88.95	91.46	2.51
4	99.50	99.50	0
5	99.49	100	0.51
6	99.00	99.5	0.50
7	99.00	100	1.00
8	100	100	-
9	100	100	-
10	100	100	-
11	98.62	99.08	0.46
12	72.25	72.25	0
13	46.36	47.73	1.37

ResNet-152 which performs best in final accuracy is used to test the generalization. From Table 3, the effect of a parallel mechanism can be identified clearly. Most subjects have an increase in the

range from 0.46% to 2.70%, due to the spectral value of different trails from different subjects having different ranges.

## 6. Conclusions and Challenges

The use of deep neural networks is widespread, while the application of frontier neural networks in EEG-based cognitive workload classification has been delayed for a certain period. One main factor which limits the use of neural networks in EEG-based classification is the lack of experimental data. In this study, we introduce the traditional methods used to dealing with EEG data, the recent works by Bashivan in the workload classification area and use his public dataset recording 2670 trails of experimental data. Compared to other public datasets only recording trails no more than a few hundred, the dataset can meet the training needs of neural networks. In this study, the method of using spectral maps is improved and a parallel mechanism of spectral feature-enhanced maps is proposed addressing the problems with compressed structural information.

The final results demonstrate the effects of adding spectral feature-enhanced maps into a network depends on individuals. In the test of generalization among 13 subjects, the maximum increase ratio is 2.7%. The final accuracy test has a mean value of 93.71% with a standard deviation of 0.0027.

Actually, the monitoring of cognitive load is still in the laboratory stage. This is partly because EEG raw signals are extremely large and when using mobile EEG facilities there is usually a delay. Some breakthroughs in wireless sensor networks, cloud computing and mobile edge computing (MEC) can manage these problems. Aloqaily et al. built a generalization of quality-of-experience (QoE) design in the cloud computing area and eventually fulfilled the service requirements by compromising between delay, service cost and information revealed to the TTP [40]. Baker et al. proposed a GreeDi-based reactive routing protocol aimed at selecting the most efficient route in terms of energy consumption between two nodes in VANETs [41]. Otoum et al. built a hierarchical trust-based WSN monitoring model for the smart grid equipment in order to detect Black-Hole (B-H) attacks and the proposed Adaptively Supervised and Clustered Hybrid IDS (ASCH-IDS) for wirelessly connected sensor performs at 98.9% detection rate and approximately 99.80% overall accuracy to detect known and unknown malicious behavior in a sensor network [42,43]. Meneguette et al. proposed an effective mobile content delivery solution for network delivery and increased content availability without compromising network overhead, regardless of traffic conditions and road networks [44]. Ridhawi et al. introduced a fog-to-fog (F2F) data caching and selection method, which allows IoT devices to retrieve data in a faster and more efficient way. Their presented solution provides guaranteed and fast delivery of the requested cloud composite services to end users while sustaining QoS requirements and load balancing among edge and mobile nodes [45]. As the edge computing paradigm began to take precedence, a mobile device cloud (MDC) formed at the edge based on idle intra-device resources emerged [46,47]. All these novel breakthroughs provide technical support for a wider usage of EEG-based monitoring, which is our future work.

**Author Contributions:** Y.Z. and Y.S. conceived the main idea, designed the main algorithm and wrote the manuscript. Y.S. designed the main experiments under the supervision of Y.Z., and the experimental results were analyzed by Y.Z.

**Funding:** This research was funded by [the Fundamental Research Funding for the Central Universities of Ministry of Education of China] grant number [18D110408] and [the Special Project Funding for the Shanghai Municipal Commission of Economy and Information Civil-Military Inoculation Project “Big Data Management System of UAVs”] grant number [JMRH -2018-1042]. And the APC was funded by [the Fundamental Research Funding for the Central Universities of Ministry of Education of China].

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Savelainen, A. An Introduction to EEG Artifacts. Available online: <https://pdfs.semanticscholar.org/5cd2/3372a19187f7dfb61c1443879f392473e250.pdf> (accessed on 18 November 2018).
2. Koelstra, S.; Muhl, C.; Soleymani, M.; Lee, J.S.; Yazdani, A.; Ebrahimi, T.; Pun, T.; Nijolt, A.; Patras, I. DEAP: A Database for Emotion Analysis; Using Physiological Signals. *IEEE Trans. Affect. Comput.* **2012**, *3*, 18–31. [[CrossRef](#)]
3. Pereira, J.; Ofner, P.; Schwarz, A.; Schwarz, A.; Muller-Putz, G.R.M. EEG neural correlates of goal-directed movement intention. *Neuroimage* **2017**, *149*, 129–140. [[CrossRef](#)] [[PubMed](#)]
4. Joyce, C.A.; Gorodnitsky, I.F.; Kutas, M. Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. *Psychophysiology* **2010**, *41*, 313–325. [[CrossRef](#)] [[PubMed](#)]
5. Wang, F.; Zhang, X.; Fu, R.; Sun, G. Study of the Home-Auxiliary Robot Based on BCI. *Sensors* **2018**, *18*, 1779. [[CrossRef](#)] [[PubMed](#)]
6. Palankar, M.; De Laurentis, K.J.; Alqasemi, R.; Veras, E.; Dubey, R.; Arbel, Y.; Donchin, E. Control of a 9-DoF Wheelchair-mounted robotic arm system using a P300 Brain Computer Interface: Initial experiments. In Proceedings of the IEEE International Conference on Robotics and Biomimetics, Bangkok, Thailand, 22–25 February 2009; pp. 348–353.
7. Kong, W.; Wang, L.; Zhang, J.; Zhao, Q.; Sun, J. The Dynamic EEG Microstates in Mental Rotation. *Sensors* **2018**, *18*, 2920. [[CrossRef](#)] [[PubMed](#)]
8. Mohamed, Z.; El Halaby, M.; Said, T.; Shawky, D.; Badawi, A. Characterizing Focused Attention and Working Memory Using EEG. *Sensors* **2018**, *18*, 3743. [[CrossRef](#)] [[PubMed](#)]
9. Roslan, N.S.; Izhar, L.I.; Faye, I.; Saad, M.N.M.; Sivapalan, S.; Rahman, M.A. Review of EEG and ERP studies of extraversion personality for baseline and cognitive tasks. *Pers. Individ. Differ.* **2017**, *119*, 323–332. [[CrossRef](#)]
10. Wang, S.; Gwizdka, J.; Chaovalitwongse, W.A. Using wireless EEG signals to assess memory workload in the *n*-back task. *IEEE Trans. Hum.-Mach. Syst.* **2016**, *46*, 424–435. [[CrossRef](#)]
11. Wang, H.; Zheng, W. Local temporal common spatial patterns for robust single-trial EEG classification. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2008**, *16*, 131–139. [[CrossRef](#)]
12. Ramoser, H.; Muller-Gerking, J.; Pfurtscheller, G. Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Trans. Rehabil. Eng.* **2000**, *8*, 441–446. [[CrossRef](#)]
13. Sun, S.; Zhang, C. Adaptive feature extraction for EEG signal classification. *Med. Biol. Eng. Comput.* **2006**, *44*, 931–935. [[CrossRef](#)] [[PubMed](#)]
14. Blankertz, B.; Tomioka, R.; Lemm, S.; Kawanabe, M.; Muller, M. Optimizing spatial filters for robust EEG single-trial analysis. *IEEE Signal Process. Mag.* **2008**, *25*, 41–56. [[CrossRef](#)]
15. Novi, Q.; Guan, C.; Dat, T.H.; Xue, P. Sub-band common spatial pattern (SBCSP) for brain-computer interface. In Proceedings of the 3rd International IEEE/EMBS Conference on Neural Engineering, Kohala Coast, HI, USA, 2–5 May 2007; Volume 2007, pp. 204–207.
16. Ang, K.K.; Chin, Z.Y.; Zhang, H.H.; Guan, C.T. Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface. In Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN) 2008, Hong Kong, 1–8 June 2008; Volume 2008, pp. 2390–2397.
17. Thomas, K.P.; Guan, C.; Lau, C.T.; Vinod, A.P.; Ang, K.K. A new discriminative common spatial pattern method for motor imagery brain-computer interfaces. *IEEE Trans. Biomed. Eng.* **2009**, *56*, 2730–2733. [[CrossRef](#)] [[PubMed](#)]
18. Shenoy, H.V.; Vinod, A.P.; Guan, C. Shrinkage estimator based regularization for EEG motor imagery classification. In Proceedings of the Communications and Signal Processing (ICICS) on IEEE, Singapore, 2–4 December 2015; pp. 1–5.
19. Kim, Y.; Ryu, J.; Kim, K.K.; Took, C.C.; Mandic, D.P.; Park, C. Motor imagery classification using mu and beta rhythms of EEG with strong uncorrelating transform based complex common spatial patterns. *Comput. Intel. Neurosci.* **2016**, *2016*, 1–13. [[CrossRef](#)] [[PubMed](#)]
20. Major, T.C.; Conrad, J.M. The effects of pre-filtering and individualizing components for electroencephalography neural network classification. In Proceedings of the SoutheastCon IEEE, Charlotte, NC, USA, 30 March–2 April 2017.

21. Pinheiro, O.R.; Alves, L.R.G.; Romero, M.F.M.; Souza, J.R.D. Wheelchair simulator game for training people with severe disabilities. In Proceedings of the International Conference on Technology & Innovation in Sports IEEE, Vila Real, Portugal, 1–3 December 2016.
22. Putze, F.; Jarvis, J.P.; Schultz, T. Multimodal Recognition of Cognitive Workload for Multitasking in the Car. In Proceedings of the 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 3748–3751.
23. Heger, D.; Putze, F.; Schultz, T. Online workload recognition from EEG data during cognitive tests and human-machine interaction. *Lect. Notes Comput. Sci.* **2010**, *6359*, 410–417.
24. Bashivan, P.; Rish, I.; Yeasin, M.; Codella, N. Learning representations from EEG with deep recurrent convolutional neural networks. *arXiv*, 2016; arXiv:1511.06448.
25. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016.
26. Huang, G.; Liu, Z.; Weinberger, K. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; Volume 2016, pp. 2261–2269.
27. Sternberg, S. The discovery of processing stages: Extensions of Donders' method. *Acta Psychol.* **1969**, *30*, 276–315. [[CrossRef](#)]
28. Sternberg, S. Memory-scanning: Mental processes revealed by reaction-time experiments. *Am. Sci.* **1969**, *57*, 421–457.
29. Bashivan, P.; Yeasin, M.; Bidelman, G.M. Single trial prediction of normal and excessive cognitive load through EEG feature fusion. In Proceedings of the Signal Processing in Medicine & Biology Symposium IEEE, Philadelphia, PA, USA, 12 December 2015; Volume 2015, pp. 1–5.
30. Michels, L.; Bucher, K.; Lüchinger, R.; Klaver, P.; Martin, E.; Jeanmonod, D.; Brandeis, D. Simultaneous EEG-fMRI during a working memory task: Modulations in low and high frequency bands. *PLoS ONE* **2010**, *5*, e10298. [[CrossRef](#)]
31. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; Volume 1, pp. 1097–1105.
32. Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [[CrossRef](#)]
33. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. ImageNet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, USA, 20–25 July 2009; Volume 2009, pp. 248–255.
34. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv*, 2014; arXiv:1409.1556.
35. Zeng, X.; Martinez, T.R. Distribution-balanced stratified cross-validation for accuracy estimation. *J. Exp. Theor. Artif. Intell.* **2000**, *12*, 12. [[CrossRef](#)]
36. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
37. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2014**, *115*, 211–252. [[CrossRef](#)]
38. Pan, S.J.; Tsang, I.W.; Kwok, J.T.; Yang, Q. Domain adaptation via transfer component analysis. *IEEE Trans. Neural Netw.* **2011**, *22*, 199–210. [[CrossRef](#)] [[PubMed](#)]
39. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.
40. Aloqaily, M.; Kantarci, B.; Mouftah, H.T.L. A Generalized Framework for Quality of Experience (QoE)-Based Provisioning in a Vehicular Cloud. In Proceedings of the 2015 IEEE International Conference on Ubiquitous Wireless Broadband (ICUWB), Montreal, QC, Canada, 4–7 October 2015; pp. 1–5.
41. Baker, T.; García-Campos, J.M.; Gutiérrez Reina, D.; Toral, S.; Tawfik, H.; Al-Jumeily, D.; Hussain, A. GreeAODV: An Energy Efficient Routing Protocol for Vehicular Ad Hoc Networks. In *Intelligent Computing Methodologies*; Springer: Cham, Switzerland, 2018; pp. 670–681.

42. Otoum, S.; Kantarci, B.; Mouftah, H.T. Hierarchical trust-based black-hole detection in WSN-based smart grid monitoring. In Proceedings of the 2017 IEEE International Conference on Communications (ICC), Paris, France, 21–25 May 2017.
43. Otoum, S.; Kantarci, B.; Mouftah, H.T. Adaptively Supervised and Intrusion-Aware Data Aggregation for Wireless Sensor Clusters in Critical Infrastructures. In Proceedings of the 2018 IEEE International Conference on Communications (ICC), Kansas City, MO, USA, 20–24 May 2018.
44. Meneguetto, R.; Boukerche, A.; Silva, F.; Villas, L.; Ruiz, L.; Antonio, A.F.L. A Novel Self-Adaptive Content Delivery Protocol for Vehicular Networks. *Ad Hoc Netw.* **2018**, *73*, 1–13. [[CrossRef](#)]
45. Ridhawi, I.A.; Mostafa, N.; Kotb, Y.; Aloqaily, M.; Abualhaol, I. Data caching and selection in 5G networks using F2F communication. In Proceedings of the 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Montreal, QC, Canada, 8–13 October 2017.
46. Ismaeel, A.R.; Moayad, A.; Yehia, K.; Yousif, A.; Yaser, J. A collaborative mobile edge computing and user solution for service composition in 5G systems. *Trans. Emerg. Telecommun. Technol.* **2018**, *29*, e3446.
47. Balasubramanian, V.; Aloqaily, M.; Zaman, F.; Jararweh, Y. Exploring Computing at the Edge: A Multi-Interface System Architecture Enabled Mobile Device Cloud. In Proceedings of the 2018 IEEE 7th International Conference on Cloud Networking (CloudNet), Tokyo, Japan, 22–24 October 2018.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).