Visual Fatigue Assessment Based on Multi-task Learning

Danli Wang, Xueyu Wang, Yaguang Song, Qian Xing, and Nan Zheng
State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China
E-mail: danliwang2009@gmail.com

Abstract. *In recent years, with the rapid development of stereoscopic display technology, its applications have become increasingly popular in many fields, and, meanwhile, the number of audiences is also growing. The problem of visual fatigue is becoming more and more prominent. Visual fatigue is mainly caused by vergence–accommodation conflicts. An evaluation experiment was conducted, and the electroencephalogram (EEG) data of the subjects were collected when they were watching stereoscopic content, and then the stereoscopic fatigue state of the subjects during the viewing process was analyzed. As deep learning is proved to be an effective end-to-end learning method and multi-task learning can alleviate the problem of lacking annotated data, the authors establish a user visual fatigue assessment model based on EEG by using multi-task learning, which can effectively obtain the user's visual fatigue status, so as to make the comfort designs to avoid the harm caused by user's visual fatigue.* *© 2019 Society for Imaging Science and Technology.*
[DOI: 10.2352/J.ImagingSci.Technol.2019.63.6.060414]

1. INTRODUCTION

In 2009, the three-dimensional (3D) movie Avatar achieved huge success and swept the world, making stereoscopic display technology enter people's life. Later, with the rapid development of 3D display technology, people can watch 3D stereo content through 3DTV and other devices. Compared with ordinary two-dimensional (2D) display technology, 3D display can provide a more realistic visual experience [1]. However, because of the conflict between human visual system and the imaging principle of 3D display technology, continuous viewing of 3D content will cause various visual discomforts, such as dry eyes, blurred vision, and even feeling giddy and dazzled [2]. In order to relieve the discomfort, we should evaluate the symptoms by classification and grading evaluation first, and then take corresponding technical improvement according to different symptom types; so it is meaningful to evaluate visual fatigue. In recent years, visual fatigue assessment has attracted the attention of many researchers [3]. The common evaluation methods are subjective evaluation method and objective evaluation method.

Subjective evaluation methods are mature now, which usually use a questionnaire to obtain the fatigue degree of the viewers. The questionnaire method is usually based on a large number of problems and gradually determines the degree of visual fatigue of users from various aspects. As a widely used method in many fields, subjective evaluation method is easy to design and implement. However, it essentially depends on the evaluation of users' self-perception, which is largely influenced by individual differences and psychological factors [3].

Because of the above problems of subjective evaluation, people have also explored the objective evaluation method of visual fatigue. Commonly used objective indicators are near point distance [4], flash fusion frequency, reaction time [5], pupil diameter, heart rate, and so on [6, 7]. Recently, electroencephalogram (EEG) is considered to be effective and reliable physiological signals. Neuronal activity in the brain can reflect the fatigue state after continuously viewing stereoscopic display content [8]. Therefore, the study of visual fatigue assessment based on EEG signals has attracted much attention. Most of the related studies focus on comparing the changes of some indicators before and after stereoscopic display. There are few works that model visual fatigue assessment based on EEG [9].

At present, some researches begin to use machine learning to extract characteristic information of EEG and do classification tasks. Visual fatigue assessment based on traditional machine learning methods usually includes several steps, such as spatial filtering, feature extraction, and classification. The whole process is complex and requires sufficient domain knowledge. Also, manually extracting features is time consuming and has good generalization performance in some tasks.

In the past two years, some researchers have applied deep learning methods to EEG classification tasks. Schirrmeister et al. (2017) explored the structure of convolution neural networks for motor imagery tasks and proposed Deep ConvNet and Shallow ConvNet, both of which outperformed the traditional Filter Bank Common Spatial Patterns (FBCSP). Lawhern et al. (2018) proposed a general convolution network EEGNet, which achieved similar results compared with traditional methods in many different EEG classification tasks. Although these methods had some innovations, the improvements were not obvious, and the development of deep learning in the field of EEG classification had been limited. This limitation comes from the problem of deep learning, which requires a lot of annotated data. However, the lack of labeling data limits the further development of in-depth learning methods because

Received July 15, 2019; accepted for publication Nov. 7, 2019; published online Jan. 13, 2020. Associate Editor: Jia-Shing Sheu.
1062-3701/2019/63(6)/060414/8/$25.00

of the lack of large-scale labeling data in EEG classification tasks.

Multi-task learning is a learning paradigm of machine learning. Its purpose is to make full use of the information contained in multiple related tasks to improve the generalization performance of the model on all tasks [10]. For the problem of insufficient training data, multi-task learning is a good solution [11]. Considering the limited EEG data, we introduce the concept of multi-task learning into visual fatigue assessment and propose a deep learning model based on multi-task learning for EEG signal classification.

The contributions of our work are as follows: we proposed a user visual fatigue assessment model based on EEG by using multi-task learning. By comparing with other deep learning methods, the results demonstrate that the proposed multi-task learning approach outperforms the state-of-the-art approaches.

2. CLASSIFICATION MODEL OF EEG SIGNALS BASED ON MULTI-TASK LEARNING

2.1 *Multi-Task Learning*

Multi-task learning is a learning paradigm of machine learning. Its purpose is to make full use of the information contained in multiple related tasks to improve the generalization performance of the model on all tasks.

Generally speaking, for most machine learning tasks, training a sufficiently accurate classification model usually requires a large amount of annotated data. However, in some applications, such as medical image analysis, EEG classification, etc., it is hard to meet such a condition, as the data acquisition process is complex and labeling requires a lot of work. In the case of limited training data, it is difficult to train a shallow model, let alone a more complex deep model. For insufficient data, multi-task learning is a good solution when there are several related tasks.

Multi-task learning usually involves multiple tasks, which are also general learning tasks, such as supervised learning (classification or regression), unsupervised learning, and so on. These tasks are interrelated as a whole or at least in part. In this setting, it is found that joint training of these tasks can significantly improve the performance of the model compared with individual training of each task. This discovery directly leads to the birth of multi-task learning. The fundamental purpose of multi-task learning is to improve the generalization ability of the model by utilizing the correlation between multiple tasks.

2.2 *Multi-Task Learning Model Structure for EEG Signal Classification*

In the case of insufficient data, multi-task learning is an effective solution. Thus, in this article, we use multi-task learning to solve the problem of limited data. Deep learning has been proved to be effective in some EEG signal classification tasks; so this article uses multi-task learning and deep learning to solve the problem of insufficient data.

The model consists of three modules: presentation learning module, classification module, and reconstruction module, as shown in Figure 1. The following is the flow of the whole framework. First, the learning module extracts the features from the input EEG signals. Here, the features are called intermediate shared features, which are fed into the classification module and the reconstruction module, respectively, to complete the classification and reconstruction tasks. The three modules are trained simultaneously and optimized jointly in an end-to-end manner. As a bridge, the middle-level shared feature connects the classification module and the reconstruction module. Through the interaction and mutual promotion of the classification and reconstruction module, the intermediate feature retains the ability to facilitate classification and reconstruction at the same time [12]. It improves the generalization performance of the model on a single task and also improves the classification task with limited data. The previous definition and mathematical symbols are also applicable in this model. Here we will introduce three modules and the training methods one by one.

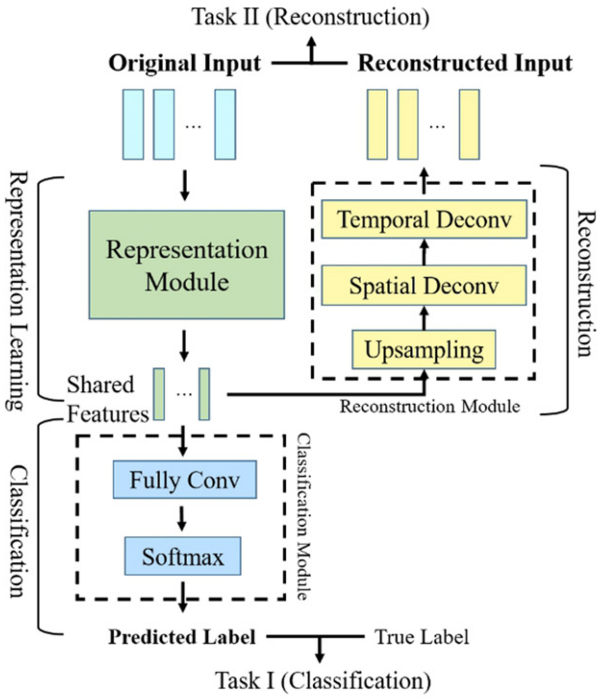

Figure 1. Overall architecture of the multi-task learning model.

2.2.1 *Representation Learning Module*

Representation learning module extracts features from the original EEG signal. The representation learning module consists of spatio-temporal convolution module, pooling layer, and batch normalization layer. The following is a brief introduction to the network structure representing the learning module:

(1) Temporal convolution layer. Convolution kernels are one-dimensional convolution kernels rather than two-dimensional convolutions commonly used in general image tasks. This layer processes the input along the

time dimension, thus compressing data to obtain a more compact structure of time dimension. The number of input channels is 1, corresponding to the number of channels of the EEG signal. The number of output channels is 40, and the convolution kernel size is (25,1).

(2) Spatial convolution layer. In order to extract the features of spatial dimension, the model also uses one-dimensional convolution. After the processing of the spatial convolution layer, the width of output feature becomes 1. As same as the time dimension convolution layer, the number of input channels is 40, the number of output channels is 40, and the size of convolution kernel is (1,30).

(3) Batch normalization layer and non-linear layer. The batch normalization layer is used to standardize the data and accelerate the convergence of the model. The Rectified Linear Units (ReLU) function is used in the non-linear layer.

(4) Average pooling layer and non-linear unit. The convolution layer of the time dimension compresses along the height of the input matrix, making the time dimension more compact. The pooling layer is used to aggregate the features of time dimension and combine some low-level features into high-level features, so as to facilitate the subsequent classification and reconstruction. The size of pooling is (75,1). The log function is used as the non-linear unit.

(5) Dropout layer. This layer randomly discards part of the input features with a certain probability to reduce the risk of overfitting. The output feature of this layer is called the intermediate shared feature.

2.2.2 *Classification Module*

After the processing of the presentation learning module, the shared features are sent to the classification module, which includes a fully convolutional layer and a softmax layer. The detailed network structure is as follows:

(1) Fully convolutional layer. According to the size of intermediate features, the appropriate size of convolution kernel is selected so that the output feature size is [1,1]. According to the number of categories of the classification task, the corresponding size of convolution kernels is determined, and the output features are the corresponding activation values of each category. The number of input channels is 40. The number of output channels is 3, corresponding to three fatigue levels, respectively, and the size of convolution kernel is (19,1).

(2) Softmax layer. The output of the classification task is generally a probability distribution, which corresponds to the probability of each category. The activation value of each category is obtained at the upper level and then the corresponding probability is obtained using softmax function.

For classification tasks, this article uses cross-entropy loss to measure the performance of the model. The specific calculation is as follows:

$$\text{Loss}_{CE} = \frac{1}{N}\sum_{k=1}^{N} \text{loss}(y^k, \hat{y}^k), \tag{1}$$

where N represents the number of samples, $\hat{y}^k$ corresponds to the true category of the kth sample, y^k is the prediction of the model for the kth sample, $\text{loss}(y^k, \hat{y}^k)$ represents the cross-entropy loss of the kth sample, Loss_{CE} and represents the average cross-entropy loss of the total sample.

2.2.3 *Reconstruction Module*

Shared features are fed into the reconstruction module to reconstruct the original input. Here, the representation learning module can be regarded as an encoder, while the reconstruction module can be regarded as a decoder. Therefore, the encoder–decoder structure constitutes an autoencoder. For the decoding process, the model uses deconvolution, also known as transposition convolution, to decode the input intermediate features. Before the continuous deconvolution operation, the upsampling is applied to the intermediate features. The following is the detailed structure of the reconstruction module:

(1) Upsampling layer. This layer acts as the mirror operation of the pooling layer, interpolates the intermediate features, and restores the feature size to the size before pooling. This layer uses bilinear interpolation, and the output feature size is (351,1).

(2) Deconvolution (spatial dimension). As a mirror operation to spatial convolution in the representation learning module, the convolution kernel of this layer is also one-dimensional. After the processing of this layer, the size of output features is restored to the size before the spatial convolution. The number of input channels is 40, the number of output channels is 40, and the convolution kernel size is (1,30).

(3) Deconvolution (time dimension). As a mirror operation of the temporal convolution in the representation learning module, this layer also uses one-dimensional convolution. After the processing of this layer, the output size is restored to the same size as that of the original input. The number of input channels is 40, the number of output channels is 1, and the convolution kernel size is (25,1).

For the reconstruction task, the mean square error loss is used to measure the performance of the model. The specific calculation process is as follows:

$$\text{Loss}_{MSE} = \frac{1}{N}\sum_{k=1}^{N} \|X^k - \hat{X}^k\|^2. \tag{2}$$

Among them, N is the total number of samples, X^k is the original EEG input, $\hat{X}^k$ is the reconstruction input, and Loss_{MSE} is the reconstruction loss corresponding to the K sample.

2.3 *Multi-Task Learning Model Training*

The network structure of the three modules in the multi-task learning framework is described in detail above. The training methods of multi-task learning are described below. From the introduction above, we can see that the model consists of two tasks, supervised learning task and unsupervised learning task. For this model, a common way is two-stage training method, that is, pre-training with encoder–decoder structure, and then supervised learning. Instead of using this method, this article chooses to optimize the presentation learning module, classification module, and re-modeling module in an end-to-end manner. In principle, after joint training, the shared intermediate features will have the ability to reconstruct and classify at the same time, that is, the generalization ability of features is stronger, which makes the two tasks mutually reinforcing. Here, θ is used to represent all the parameters of the model; so the total loss function can be recorded as

$$\mathcal{L}(\Theta) = \text{Loss}_{CE} + \alpha \cdot \text{Loss}_{MSE} + \lambda \|\Theta\|^2. \tag{3}$$

Among them, Loss_{CE} denotes the loss of classification tasks, i.e., cross-entropy loss, and optimizes the model by supervised learning; Loss_{MSE} denotes the loss of reconstruction tasks, i.e., mean square error loss, and optimizes the model by unsupervised learning; and α denotes the ratio of reconstruction losses to classification losses, i.e., the relative importance of two tasks. For simplicity, α is set to a fixed value. Finally, λ is the coefficient for regularization in order to reduce the risk of overfitting.

3. EXPERIMENTAL DESIGN

3.1 *Experimental Purpose*

Deep learning has been proved to be effective in some EEG signal classification tasks. However, deep learning requires large-scale labeling data, but there is no large-scale dataset for evaluating visual fatigue using EEG. Therefore, this article proposes a multi-task learning framework for visual fatigue assessment. The evaluation experiment is based on two kinds of EEG classification tasks (single subject and multiple subjects). The purpose of the experiment is (1) to verify that the lack of labeled data really limits the further improvement of the performance of deep learning methods and (2) to verify that multi-task learning model can further improve the classification accuracy of deep learning model under the condition of limited label data.

3.2 *Dataset for Visual Fatigue Assessment*

3.2.1 *Data Acquisition*

A total of 20 participants, including 4 females, are involved in the visual fatigue assessment experiment. Their average age is 23 years old, ranging from 22 to 25 years old. All participants have been notified in advance that alcohol and other irritating drinks or food should be avoided within 24 hours before the experiment and that adequate sleep should be guaranteed for 8 hours. Before beginning the experiment, the visual acuity of all subjects is normal and the stereo parallax was less than or equal to 200. It meets the experimental requirements.

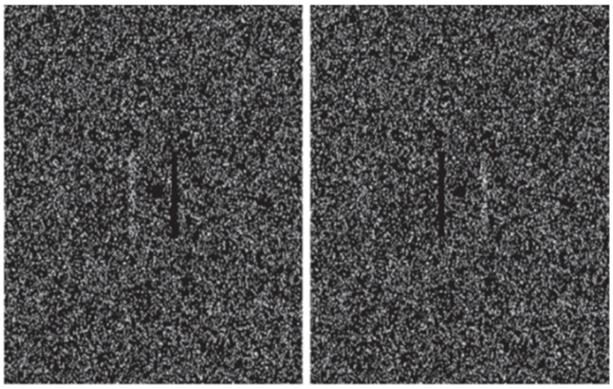

Figure 2. Left and right view of visual stimulus.

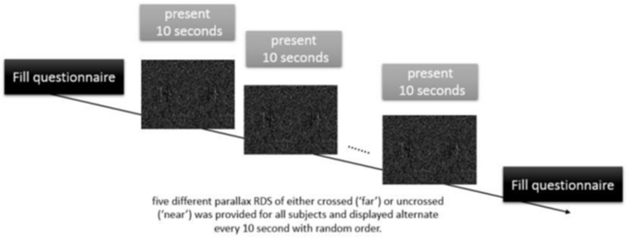
present
10 seconds
present
10 seconds
present
10 seconds
Fill questionnaire
Fill questionnaire
five different parallax RDS of either crossed ('far') or uncrossed ('near') was provided for all subjects and displayed alternate every 10 second with random order.

Figure 3. The experimental process of visual fatigue assessment.

Random dot stereogram (RDS), as a visual stimulus, is used to induce subjects to enter the visual fatigue state. The advantage of RDS is that it eliminates the influence of plot and other differences in visual stimulus. The random point stereogram used in the experiment includes a left view and a right view. As shown in Figure 2, the random point stereogram is generated by Unity 3D (Unity Technologies, USA). Five parallax settings (0°, 0.5°, –0.5°, 1.0°, and −1.0°) were used in this experiment.

The experiment is conducted in an appropriate environment to ensure that the subjects are not affected by external factors. The whole experiment is divided into six sections, which lasts about 30 minutes, as shown in Figure 3. Before beginning the experiment, all subjects have a 5-minute break to adjust their state to the best. Then five parallax maps of different degrees played in random order, each lasting 2 seconds, and each level of parallax maps appears about 15 times in each section. At the end of each section, participants take a 10-second break. During this period, participants need to choose the corresponding fatigue level according to their own state.

NeuroScan system (compumedics, Australia) is used to record the signals of 34 channels at 500 Hz. Thirty of them (Fp1, Fp2, F3, F4, F7, F8, Fz, FC3, FC4, FT7, FT8, FCZ, C3, C4, T3, T4, CZ, CP3, CP4, TP7, TP8, CPZ, P3, P4, T5, T6, PZ, O1, O2, and Oz) are used for further experimental analysis, as shown in Figure 4.

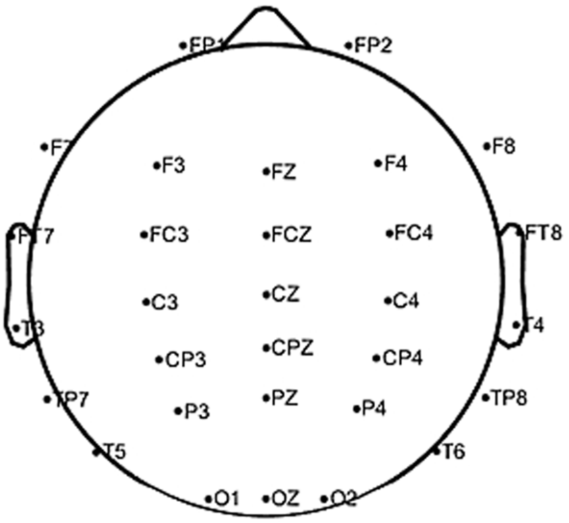

Figure 4. Distribution of EEG electrodes.

3.2.2 *Data Preprocessing*

This article mainly discusses EEG data. The latter data analysis refers to the analysis and processing of EEG data. The data preprocessing tool is EEGLAB of MATLAB. The following is the pretreatment process for the visual fatigue assessment dataset.

First, the collected signal is sampled down to 250 Hz, then the signal below 1 Hz is filtered by high-pass filter to eliminate baseline drift, and then the signal frequency is controlled at 1–40 Hz by low-pass filter, which eliminates the high-frequency information that has little correlation with the experiment and reduces the computational load of the subsequent experiment. The collected EEG signals include not only the information of brain activity needed for modeling but also other noise disturbances, such as eye movement and electro-oculogram (EOG). These artifacts exist in EEG signals in a specific mode; so independent component analysis is used to separate and remove these artifacts.

Next, the collected EEG signals are divided into equal length data segments according to a certain time step (10 seconds). After further noise removal based on the subjective evaluation of the subjects, the annotated dataset for visual fatigue assessment is obtained.

Due to the influence of limited data scale on deep learning methods and the validity of multi-task learning, the experiment contains two training paradigms: one is based on the data of a single subject for training and prediction; the other is based on the data of multiple subjects for training, and single subject for prediction. In order to ensure the quality of single-subject data, the experimental dataset is screened, eliminating the subjects with class missing and too little single-category data. The final dataset, which is used in the classification experiment, consisted of 11 subjects, each of whom contained three fatigue levels. The statistical information of each subject in the dataset is as follows (Table I).

To show the effectiveness of our model, the state-of-the-art methods on BCI Competition IV dataset 2a are chosen for comparison. The baseline methods are listed as follows:

(1) Filter Bank Common Spatial Patterns [13]: It is designed to extract band power features of EEG. A classifier is trained to predict labels based on the features. FBCSP is an extension of the traditional common spatial patterns (CSP) algorithm, and it is the best traditional method in motor imagery task. Comparing other deep learning models with FBCSP, we can verify the effectiveness of deep learning method in EEG signal classification task.
(2) Shallow ConvNet [14]: Inspired by FBCSP algorithm, Shallow ConvNet extracts features in a similar way. But Shallow ConvNet uses convolutional neural network to do all the computations and are optimized in an end-to-end manner. Shallow ConvNet is a state-of-the-art method of deep learning model proposed for EEG classification task in recent years. Compared with it, we can verify the effectiveness of multi-task learning model.
(3) Deep ConvNet [14]: It has four convolution-pooling blocks and is much deeper than Shallow ConvNet. The purpose of introducing Deep ConvNet as the baseline is to explore the effect of increasing the complexity of the model when the training data is insufficient, that is, whether the limited data is the bottleneck of the further improvement of the model for the deep learning method.
(4) EEGNet [15]: It has two convolution-pooling blocks. The difference between EEGNet and ConvNets introduced above is that EEGNet uses depthwise and separable convolution. EEGNet has excellent performance in many classification tasks based on EEG signals, showing good generalization. In order to verify the ability of generalization of multi-task learning model, EEGNet is a good baseline method.

3.3 *Evaluation Metric*

The overall accuracy for each subject is computed and the average accuracy for each method is reported. The overall accuracy is calculated as follows:

$$\text{accuracy} = \frac{\sum_{c=1}^{c} TP_c}{N}, \tag{4}$$

where TP_c is the number of the true positive samples of class c, C is the number of classes, which is three in this experiment, and N is the number of trials.

4. RESULT ANALYSIS

Based on the above settings, this section analyzes model training and prediction based on single subject and multi-task and compares the results of different models and current methods. In order to further verify the effectiveness of the multi-task learning framework, ablation experiments are also designed.

4.1 *Result of Single Subject Training Paradigm*

Table II shows the performance of different methods on visual fatigue assessment datasets. According to the average

Table I. The possible values of the three factors.

Fatigue	Subject										
level	1	2	3	4	5	6	7	8	9	10	11
1	205	204	191	171	423	206	213	427	421	411	213
2	410	189	384	346	209	203	207	643	208	583	215
3	618	607	567	355	639	202	200	215	631	198	427
Total	1233	1000	1142	872	1271	611	620	1285	1260	1192	855

Table II. Performance comparisons of different methods under single-subject training paradigm. Best scores are in bold.

	Accuracy % (mean std. dev.)			
Subject	Shallow ConvNet	Deep ConvNet	EEGNet	Ours
1	59.72.3	56.82.1	52.04.1	**62.61.5**
2	63.34.2	58.82.3	59.11.7	**64.81.7**
3	75.42.3	72.81.8	73.02.3	**75.74.5**
4	65.14.2	49.12.8	62.53.0	**70.13.5**
5	70.92.6	66.81.5	63.26.7	**73.21.4**
6	58.52.4	53.56.5	59.63.4	**65.53.8**
7	67.44.8	58.34.9	62.75.4	**71.23.0**
8	59.12.4	56.71.5	60.22.7	**61.11.6**
9	78.52.1	80.62.1	**83.21.9**	81.42.2
10	74.71.6	69.23.4	70.03.0	**75.50.7**
11	65.90.6	57.31.6	55.25.7	**69.24.0**
AVG	67.12.7	61.82.8	63.73.6	**70.02.5**

accuracy performance of the subjects, the multi-task learning model reaches 70.0%, which surpasses all other methods. It is nearly 3% higher than the state-of-the-art deep learning method, and the standard deviation of the multi-task learning model is smaller than other methods, which shows that the method is more stable. According to the classification accuracy of individual subjects, the multi-task learning model achieves the best results on almost all subjects' (1,2,3,4,5,6,7,8,10,11) datasets. The above results demonstrate the effectiveness of the multi-task learning model.

Deep ConvNet performs the worst, with an average classification accuracy of only 61.8%, which is much lower than the multi-task learning model and Shallow ConvNet. According to the results of a single subject, Shallow ConvNet outperforms Deep ConvNet in almost all subjects (1,2,3,4,5,6,7,8,10,11). According to the previous introduction, the design of Deep NetConv and Shallow ConvNet is similar. The difference is that Deep ConvNet has four convolution-pooling modules, which are more complex, and has more layer than Shallow ConvNet. Theoretically speaking, more complex structure and more parameters make the model perform better. However, a key prerequisite for this common sense is that sufficient annotated data is needed for training, and the experimental settings and results also validate the previous hypothesis that limited training data is indeed the limitation of further improvement of deep learning model in EEG classification tasks. The experimental results also show that the deeper and more complex models cannot improve the prediction accuracy. Because of overfitting, the results become worse.

The evaluation data of subject 1 with low classification accuracy in Table II are analyzed. It is found that the reason for the low accuracy is the misclassification of categories. For example, samples of category 2 of the subjects 1 are misclassified into category 3, which indicates that the subjects 1 do not have a good grasp of the evaluation standard, resulting in poor performance of the evaluation model.

In addition, in order to further verify that the improvement of prediction accuracy of the model really benefits from the multi-task learning framework, ablation experiments are carried out in this article. Figure 5 shows the comparison between the multi-task learning model and the single-classification model. The horizontal axis of the figure represents different subjects and the average, while the vertical axis of the figure represents the corresponding classification accuracy. According to the average accuracy of the graph, the multi-task model is 1.8% higher than the single-classification model. In addition, the multi-task learning model performs better in almost all the subjects, which proves the effectiveness of the multi-task learning framework. The above results and analysis further validate the previous hypothesis that the multi-task learning model composed of classification tasks and reconstruction tasks can further improve the classification effect of the model in the case of limited annotated data. The reason is that with the multi-task learning framework, the features extracted by the learning module have the ability of classification and reconstruction, which greatly improves the generalization ability of the model.

For the multi-task learning framework, the relative importance of the two tasks has a decisive impact on the final results. In order to further explore the relationship between weight coefficients and model classification accuracy, the following parameter experiments are carried out in this chapter. As shown in Figure 6, the horizontal axis of the figure represents the different values of the weight parameters (0–9), while the vertical axis is the classification accuracy corresponding to the different weight coefficients. Careful observation of the curve in the figure shows that when the weight coefficient is 0, the model only performs classification tasks, and in this case, the final performance is not good. With the increase of weight coefficient, the importance of reconstruction task increases gradually, and the performance of the model also increases rapidly. When the weight coefficient is about 1.5, the performance of the model achieves the best. After that, with further increase of the weight coefficient, that is, the importance of reconstruction task is increasing, and the performance of the model decreases rapidly. The main reason is that the classification accuracy is the evaluation criteria.

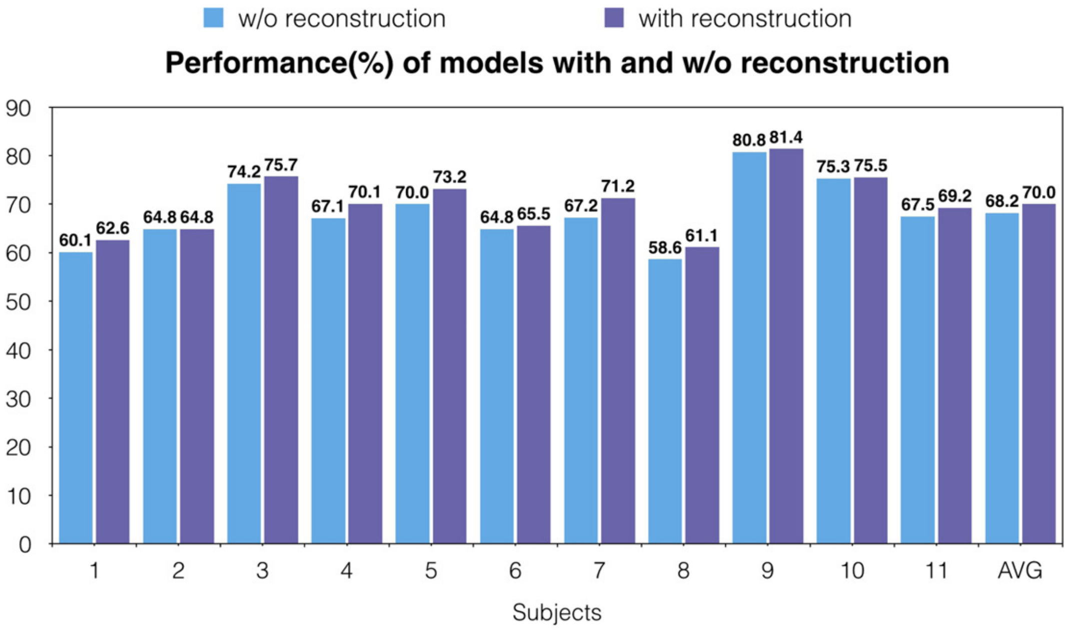

Figure 5. Performance comparison of models with and without reconstruction module. ''w/o'' is the abbreviation of ''without.''

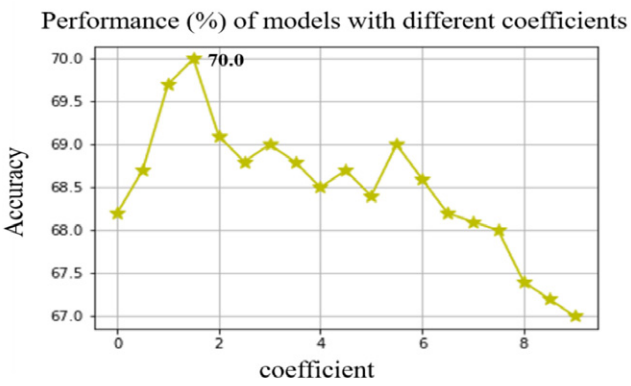

Figure 6. Performance comparison of models trained with different weight coefficients (the hyperparameter α) ranging from 0 to 9. The vertical axis corresponds to the average accuracies (%). Model achieved the best performance (70.0%) when α was about 1.5.

When the importance of reconstruction task is too great, the classification performance will naturally decrease. This result is instructive for determining the relative importance of two tasks in the experiment.

The above experimental results are based on the experimental paradigm of single subjects, that is, training and testing on the dataset of single subjects, which is also the paradigm used in most of the work. In this experimental paradigm, the label data used for training is limited, and the multi-task learning model performs best, which also proves the validity of the model.

4.2 *Result of Multi-Person Training Paradigm*

This article also adopts the paradigm of multi-subject training, which combines all the subjects' training data into a training model and then tests them on the data of a single subject. The purpose of the experiment is to explore how the performance of each model changes when the scale of annotated data expands. Consistent with previous experiments, a fivefold cross-validation is used to ensure that the experiment is not affected by randomness.

The results are shown in Table III. Under the multi-subject training mode, the overall performance of most models do not improve mainly because of the large differences between different subjects. However, the average accuracy of Deep ConvNet is higher than that of the single-subject training paradigm, and the performance of Deep ConvNet is significantly worse than that of Shallow ConvNet under the single-subject training paradigm. However, in this experiment, the deep model has surpassed the shallow model. The above results show that increasing the size of training data can indeed further improve the performance of deep model, which in turn proves that in previous experiments, limited label data is indeed the bottleneck of deep learning methods. In addition, in this experiment, the multi-task learning model still achieves the best results. The above analysis shows that the multi-task learning framework can break the limitation of insufficient data to a certain extent and improve the generalization ability of the model.

The evaluation data of subject 6 with low classification accuracy in Table III are analyzed. It is found that the reason of low classification accuracy is the misclassification of category. For example, the data of category 2 of subject 6 are misclassified into category 1, which indicates that subject 6 does not have a good grasp of the evaluation standard, resulting in poor performance of evaluation model.

5. CONCLUSIONS

In this article, for the first time, the deep learning method based on multi-task learning framework is applied to

Table III. Performance comparison of different methods under multi-subject training paradigm. Best scores are in bold.

	Accuracy % (mean std. dev.)			
Subject	Shallow ConvNet	Deep ConvNet	EEGNet	Ours
1	57.41.7	**57.73.0**	54.31.1	57.60.5
2	61.71.1	61.02.8	60.21.1	**62.91.5**
3	65.10.8	**70.22.5**	60.93.2	68.32.4
4	62.71.6	58.83.2	53.63.6	**66.42.8**
5	64.02.1	66.81.3	65.13.0	**67.71.9**
6	50.33.6	55.72.8	52.82.4	**56.32.7**
7	59.63.8	60.13.4	59.04.0	**65.02.9**
8	53.22.1	58.01.6	59.62.0	56.62.6
9	67.81.2	71.43.9	68.84.3	**71.81.8**
10	66.32.9	65.62.7	61.53.0	**66.31.6**
11	56.12.0	**59.92.9**	50.63.9	57.51.8
AVG	60.42.1	62.32.7	58.82.9	**63.32.1**

EEG signal classification tasks. We designed two training paradigms, one based on a single subject and the other based on multiple subjects. The difference between them is the size of the training dataset. By comparing with other deep learning methods, the results show that the limited annotated data is indeed the limitation of further improvement of deep learning methods. In addition, the proposed multi-task learning approach outperforms the state-of-the-art approaches and proves that the multi-task learning framework can indeed improve the generalization level of the model in the case of limited annotated data. In future, we plan to combine features from a variety of physiological signals such as ECG and user behavior data. The performance and robustness of the model can be further improved through the fusion of multi-modal data. In addition, the number of the subjects in this experiment is relatively small, which also leads to the limitation of the performance of the model. We plan to increase the number and types of subjects to better verify the performance of the model.

ACKNOWLEDGMENTS

This research is supported by the National Key Research and Development Program under Grant No. 2016YFB0401202 and the National Natural Science Foundation of China under Grant Nos. 61872363, 61672507, 61272325, 61501463, and 61562063.

REFERENCES

[1] Q. Wang and A. Wang, “Survey on stereoscopic three-dimensional display: Survey on stereoscopic three-dimensional display,” J. Comput. Appl. **3**, 579–581 (2010).

[2] T. Wang, “Evaluation of stereoscopic display viewing experience,” *University of Chinese Academy of Sciences* (2014).

[3] M. Lambooij, M. Fortuin, I. Heynderickx, and W. Ijsselsteijn, “Visual discomfort and visual fatigue of stereoscopic displays: a review,” J. Imaging Sci. Technol. **3**, 30201–1 (2009).

[4] N. Kim, J. Park, and S. Oh, “Visual clarity and comfort analysis for 3D stereoscopic imaging Contents,” Int. J. Comput. Sci. Netw. Secur. **2**, 227–231 (2011).

[5] C. J. Lin, Y. H. Hsieh, H. C. Chen, and J. C. Chen, “Visual performance and fatigue in reading vibrating numeric displays,” Displays **4**, 386–392 (2008).

[6] D. Wang, Y. Qi, T. Wang, H. Qiao, Y. Shi, and L. Zhang, “Evaluation of stereoscopic visual fatigue in experts,” J. Soc. Inf. Disp. **8**, 437–447 (2015).

[7] X. Yang, D. Wang, H. Hu, and Y. Kang, “Visual fatigue assessment and modeling based on ECG and EOG caused by 2D and 3D displays,” Sid Symp. Dig. Tech. Pap. **1**, 1237–1240 (2016).

[8] K. Hirvonen, S. Puttonen, K. Gould, J. Korpela, V. F. Koefoed, and K. Müller, “Improving the saccade peak velocity measurement for detecting fatigue,” J. Neurosci. Methods **2**, 199–206 (2010).

[9] Y. G. Song, D. L. Wang, K. Yue, and N. Zheng, “DeepFatigueNet: A model for automatic visual fatigue assessment based on raw single-channel EEG,” C. SID Symposium Digest of Technical Papers **50**, 965–968 (2019).

[10] S. Ruder, “An overview of multi-task learning in deep neural networks,” arXiv:1706.05098 (2017).

[11] Y. Zhang and Q. Yang, “A survey on multi-task learning,” arXiv:1707.08114 (2017).

[12] Y. Zhang, D. Shen, G. Wang, Z. Gan, R. Henao, and L. Carin, “Deconvolutional paragraph representation learning,” *J. Advances in Neural Information Processing Systems*, 4169–4179 (2017).

[13] K. K. Ang, Z. Y. Chin, C. Wang, C. Guan, and H. Zhang, “Filter bank common spatial pattern algorithm on BCI competition IV datasets 2a and 2b,” Front. Neurosci. **6** (2012).

[14] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiderer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, “Deep learning with convolutional neural networks for EEG decoding and visualization,” Hum. Brain Mapp. **11**, 5391–5420 (2017).

[15] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, “EEGNet: a compact convolutional network for EEG-based Brain-Computer Interfaces,” J. Neural Eng. **5**, 056013 (2018).