Emotion Recognition under Sleep Deprivation Using a Multimodal Residual LSTM Network

Le-Yan Tao[1], Bao-Liang Lu[1, 2, 3, 4, *]
[1]Center for Brain-like Computing and Machine Intelligence
Department of Computer Science and Engineering
[2]Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering
[3]Brain Science and Technology Research Center; [4]Qing Yuan Research Institute
Shanghai Jiao Tong University, Shanghai, 200240, China

***Abstract*—Emotion recognition under sleep deprivation is instructive for the study of mental disorders such as major depressive disorder. Previous studies on emotion recognition under sleep deprivation have been mainly based on psychological research techniques. In this paper, we introduce a multilayer weight-sharing multimodal residual LSTM network for emotion recognition under sleep deprivation. The advantage of our proposed method is that it allows for the combination of three different features: the electroencephalography (EEG) single-channel differential entropy (DE) features, EEG functional strength features with topological correlation connectivity, and eye movement features. The experiments under the conditions of sleep deprivation, sleep recovery and baseline are designed and conducted. The experimental results demonstrate that the proposed method significantly enhances the performance compared with the simple concatenation of the features of different modalities, and the best mean accuracies of 86.86% and 82.03% are achieved for four emotions (happiness, sadness, fear, and neutral) in subject-dependent and cross-subject emotion recognition tasks under 30 hours of sleep deprivation, respectively. The classification accuracy of the happiness emotion is obviously impaired under sleep deprivation, indicating that sleep deprivation impairs the stimulation of the happiness emotion, and one night of sleep recovery can reactivate the elicitation of the happiness emotion to the baseline level. Furthermore, we study the brain neural patterns of the four emotional states. The prefrontal area becomes less activated for the happiness emotion and sadness emotion in the gamma band under sleep deprivation, while the neural pattern of the fear emotion is highly robust with respect to sleep deprivation.**

***Index Terms*—emotion recognition, sleep deprivation, long short-term memory network, electroencephalography (EEG)**

I. Introduction

Emotion recognition under sleep deprivation is a particularly interesting field of research, because almost all psychiatric and neurological mood disorders co-occur with sleep abnormalities. Subjective self-reports of associated irritability and behavioral volatility due to sleep deprivation are available in the literature [1]. This association indicates a potential intimate interdependence between sleep conditions and emotional functioning. An emerging consensus suggests that sufficient sleep plays a vital role in the recalibration of the emotional processing of the brain [2], while sleep deprivation reduces the capacities of emotional regulation and leads to the loss of perceptual sensitivity of critical emotional information about the external environment and the internal milieu [3].

The majority of studies on emotion recognition under sleep deprivation have been based on psychological research techniques, including subjective approaches and objective approaches. Subject rating scales often utilize a sleep restriction paradigm with questionnaire emotion scales [4], in which the questionnaires require the participants to self-evaluate the elicitation level of certain types of emotions. However, these subjective approaches are unable to deal with the individual evaluation scaling differences across participants and the gap between the ground-truth emotional elicitation level and the self-evaluated emotional elicitation level. Using approaches that objectively evaluate the performance of the participants on emotion recognition with sleep restriction, several studies have investigated the effects of sleep deprivation on emotional processing. Emotional facial expression rating is a commonly used objective task. Pallesen *et al.* [5] demonstrated that the accuracy and the speed of the rating of emotional facial expressions deteriorate simultaneously following one night of sleep deprivation. Wagner *et al.* [6] reported that sleep significantly improves the accuracy of recognizing emotional facial expressions.

Emotion recognition under sleep deprivation has also been investigated at the physiological level. The functional MRI (fMRI) method has been often applied to study the internal emotional states. Yoo *et al.* [7] suggested that sleep deprivation can result in a weakening of the capacity of higher-order brain areas to regularly control the primitive threat detection systems and emotional reactivity systems. Eye movements have also been employed to study the external subconscious behaviors of emotion recognition under sleep deprivation. Franzen *et al.* [8] demonstrated that the pupillary responses to negative emotional images are more significant compared to the positive or neutral emotional images in sleep-deprived participants.

Since emotions are complex psychophysiological processes that are associated with internal emotional states and external subconscious behaviors, it is essential to take advantage of the intramodality and intermodality correlations that contain different aspects of information underlying different types of emotions. The integration of different modalities with fusion technologies has been widely applied for multimodal emotion

*Corresponding author: Bao-Liang Lu (bllu@sjtu.edu.cn)

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

recognition. The combination of auditory and visual modalities [9]. Other studies applying the integration of emotion-related physiological modalities, e.g., EEG and eye movements, that correspond to the internal and external physiological representation of emotion recognition, respectively, has been reported to outperform the former method in emotion recognition [10] [11].

To effectively extract temporal information from the EEG signals and eye movements, we adopt a multimodal residual LSTM network that learns the intramodality high-level temporal features with multiple LSTM layers for each type of features and the intermodality correlations among the three different types of features by sharing the weights across the parallel LSTM structures in the same layer [12]. This weight-sharing architecture across modalities was first proposed for speaker identification [13]. It also achieved satisfactory performance when it was implemented for multimodal emotion recognition using raw EEG signals and raw peripheral physiological signals (PPS) signals [12]. A significant performance improvement was obtained from the complementation and the competition among all of the modalities by learning shared weights across modalities.

This paper has two innovation aspects. On the one hand, we investigate the emotion recognition under sleep deprivation using deep neural networks instead of using subjective or objective psychological research approaches. On the other hand, we integrate the intramodality and intermodality correlations of EEG single-channel DE features, EEG strength features with topological correlation connectivity, and eye movement features for multimodal emotion recognition using the multimodal residual LSTM network, as illustrated in Fig. 1. The main contributions of this paper are as follows.

1) We perform experiments under sleep deprivation condition, sleep recovery condition, and baseline condition for recognizing four types of emotions (happiness, sadness, fear, and neutral emotions).
2) We adopt a multilayer weight-sharing multimodal residual LSTM network for combining the intramodality and intermodality emotional correlations of EEG single-channel DE features, EEG strength features with correlation connectivity, and eye movement features for the first time.
3) We improve the performance by using the multimodal residual LSTM network in both subject-dependent and cross-subject emotion classification tasks.
4) We investigate the emotion types and neural emotional patterns impaired by sleep deprivation.

The remainder of this paper is organized as follows. Section II provides a brief review of the related work on emotion recognition with satisfactory performance using different modalities of features. Section III describes feature extraction and the multimodal residual LSTM network that we adopt in this paper. Section IV introduces the experiment setup. Section V presents the experimental results and discussion. Finally, we summarize our work in Section VI.

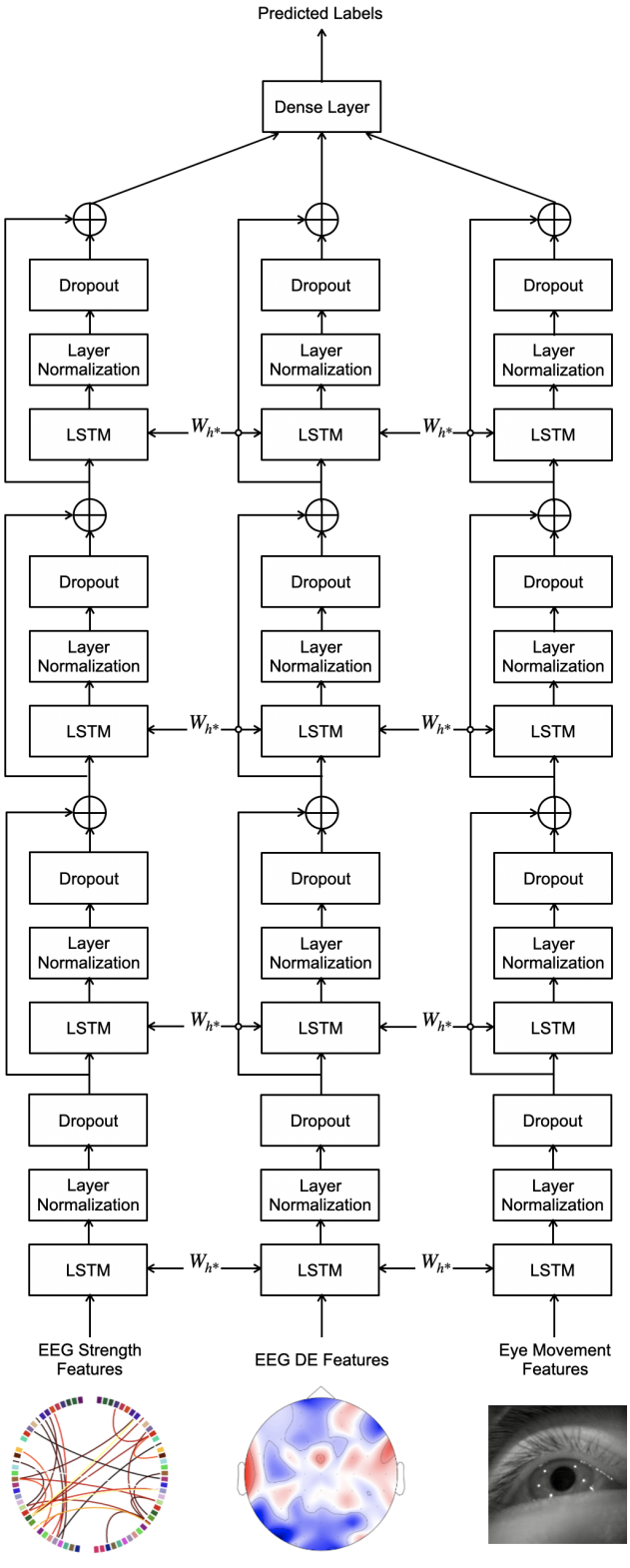


Fig. 1. Multimodal residual LSTM network

II. Related Work

In the context of mood disorders, some works using information processing technology have been proven to be efficient. Mahendran *et al.* [20] developed a stacking-based ensemble learning model using multilayer perceptron, SVM and random forest as low-level learners to diagnose major depressive disorders. Jadhav *et al.* [32] applied the decision tree classifier to screen bipolar disorder using the Mood Disorder Questionnaire.

As an essential component of studying mental disorders, emotion recognition has been performed using various modalities, such as voice, facial expressions, EEG, pupillary diameter (PD), and electrocardiograph (ECG). Among these modalities, for emotion recognition using single modality, EEG signals have been widely utilized to develop effective brain-computer interaction systems for their representation of human internal emotional and cognitive states. EEG-based emotion recognition is mainly applied using single-channel analysis. Specifically, features such as the most common used power spectral density (PSD) features and differential entropy (DE) features are independently extracted from each EEG channel in each frequency band. DE features are specifically preferred because they reflect the energy changes in the EEG signals [14] [15]. However, these single-channel-based features only reflect neural activities within a single EEG channel but fail to take advantage of the functional connectivity information among the EEG channels in different brain areas.

A few studies on EEG-based emotion recognition have focused on exploiting the brain functional connectivity among the EEG channels. Song *et al.* [16] modeled the multichannel features based on the dynamical graph convolutional neural networks. Chen *et al.* [17] and Lee *et al.* [18] directly used connectivity indices including correlation, coherence, phase synchronization, and mutual information as features without taking the brain network topology into consideration. Wu *et al.* [19] explored the emotion associated functional brain connectivity patterns by using a critical subnetwork selection approach and extracting the topological features based on the brain connectivity networks and achieved an enhancement of 3.78% by the decision-level fusion of the DE features and strength features compared with the state-of-the-art result solely using the DE feature on the SEED dataset.

Eye movement signals have been widely used in human-computer interaction (HCI) research for usability analysis and assessment because they provide an efficient and convenient method for observing user behaviors. Most previous works use eye movements to analyze the interests of users, visual search processes, and information processing. Eye movement signals enable the determination of what attracts the attention of the users and the observation of their subconscious behaviors. Eye movement signals are also essential cues for the context-aware environment that contains complementary information for emotion recognition. Some previous studies have developed effective eye movement features for emotion recognition. These studies mostly focus on pupillary responses to different emotions [10] or the combination of pupillary responses, blink, fixation and saccade information [21].

Since emotions are complicated psychophysiological phenomena associated with nonverbal cues, it is difficult to build robust emotion recognition models using only a single modality of physiological signals. In addition to the abovementioned studies that focused on a single EEG modality or a single eye movement modality, multimodal approaches have also been widely implemented for emotion recognition. Lu *et al.* [11] applied a fuzzy integral fusion strategy to combine EEG features and eye movement features on the SEED dataset. Lin *et al.* [22] transformed the EEG signals into images and extracted the hand-crafted features of other peripheral physiological signals to train a deep CNN. Zhang *et al.* [23] used group sparse canonical correlation analysis to investigate the group structure information among the EEG and eye movement features and to obtain a fusion representation of EEG and eye movement to detect anxiety. However, these works do not explicitly model the temporal correlations among the multiple modalities for emotion recognition; instead, their approaches are generally based on feature concatenation, common layers or decision ensemble. Therefore, our proposed method aims to improve the previous studies by using a deep multimodal residual LSTM network with temporal weights shared across the multiple modalities, including EEG single-channel DE features, EEG strength features with topological correlation connectivity, and eye movement features.

III. Methods

A. Feature Extraction

To investigate emotion recognition under sleep deprivation, we exploit three types of physiological features: the EEG differential entropy (DE) features, EEG strength features with correlation connectivity, and eye movement features.

a) Differential Entropy Features: EEG signals can be generally divided into five different frequency bands, namely, delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (30-50 Hz). The DE feature extraction procedure converts the EEG signals from the time domain to the frequency domain and then extracts useful information for the five frequency bands. DE features are commonly used due to their efficacy for reflecting the energy change of the EEG signals [14]. We extracted DE features from the 62-channel EEG signals in five frequency bands using short-term Fourier transforms with a 1 s nonoverlapping time window, for a feature vector with the total length of 310. The linear dynamic system approach is adopted to filter out the components of the DE features that are not associated with emotional states [24].

b) Strength Features with Correlation Connectivity: EEG strength features are derived according to critical subnetworks of the five aforementioned frequency bands. We select the critical subnetworks based on a 62×62 symmetric connectivity matrix that represents the connections between pairs of EEG channels in each frequency band for each 1-second sample. To eliminate the disturbance of the emotion-irrelevant connections, critical subnetwork selection is applied to identify

the common emotion associated connectivity patterns among the different subjects under different sleep conditions based on a tuned threshold [25]. From these matrices, we can extract the strength features with correlation connectivity that are demonstrated to be the topological features with the best emotion classification performance [19]. The linear dynamic system is also adopted for the extracted strength features for feature smoothing, and minimal redundancy maximal relevance (mRMR) [26] is employed to filter the emotion-irrelevant strength features and diminish the curse of dimensionality.

c) Eye Movement Features: Eye movement features are extracted from different detailed parameters used in the literature, such as pupil diameter, fixation and saccade [21]. The details of the features extracted from eye movements are shown in Table I.

TABLE I
DETAILS OF EXTRACTED EYE MOVEMENT FEATURES

Parameters	Extracted features
Pupil diameter (X and Y)	Mean, standard deviation and DE features in four bands: 0-0.2 Hz, 0.2-0.4 Hz, 0.4-0.6 Hz, and 0.6-1 Hz.
Pupil dispersion (X and Y)	Mean, standard deviation.
Fixation duration	Mean, standard deviation.
Saccade	Mean, standard deviation of saccade duaration and saccade amplitude. Mean, standard deviation of peak speed, average speed, peak acceleration, peak deceleration, average acceleration.
Event statistics	Fixation frequency. Average, maximum and minimum of fixation duaration. Average, maximum and minimum of pupil dispersion (X and Y). Saccade frequency. Average, maximum and minimum of saccade duration and saccade amplitude. Average saccade latency. Scanpath Length.

For both the subject-independent emotion classification tasks and the cross-subject emotion classification tasks, the recorded data from the 63-second continuous period of each emotion stimuli clip were used for extracting the EEG DE features, EEG strength features with correlation connectivity, and eye movement features.

B. Multimodal Residual LSTM Network

To exploit the combination of EEG DE features, EEG strength features with correlation connectivity, and eye movement features for emotion recognition, we adopt the multimodal residual LSTM network, as illustrated in Fig. 1.

LSTM [27] is a popular variant of recurrent neural networks and serves as the basic component of each layer of the model for its effectiveness in the extraction of temporal information from long-term biosignals [28] [29]. Temporal information is stored in the cell states c_t that propagate through time. Three data-driven gates, the forget gate f_t, the input gate i_t, and the output gate o_t, are responsible for protecting and controlling the cell state c_t.

The multimodal residual LSTM network explicitly learns the correlations among the three different types of features by sharing the weights W_{h*} across the LSTM structure in the same layer [12]. Because of the complexity of high-level temporal feature learning with explicit correlation control, we used the multimodal residual LSTM network with multiple LSTM layers for each type of feature. As depicted in Fig. 1, we adopt the multimodal residual LSTM network that consists of three 4-layer parallel LSTM structures sharing the weights W_{h*}, and each structure corresponds to the input sequences of EEG DE features, EEG strength features with correlation connectivity, and eye movement features. The formulas for the multimodal residual LSTM network, excluding the bias terms, are given as follows:

$$\tilde{c}_t^s = tanh(W_{hg} * h_{t-1}^s + W_{xg}^s * x_t^s),$$

$$f_t^s = \sigma(W_{hf} * h_{t-1}^s + W_{xf}^s * x_t^s),$$

$$i_t^s = \sigma(W_{hi} * h_{t-1}^s + W_{xi}^s * x_t^s),$$

$$o_t^s = \sigma(W_{ho} * h_{t-1}^s + W_{xo}^s * x_t^s),$$

$$c_t^s = f_t^s \odot_{t-1}^s + i_t^s \odot \tilde{c}_t^s,$$

$$h_t^s = o_t^s \odot tanh(c_t^s),$$

where the shared weights W_{h*} across the three parallel LSTM structures including W_{hg}, W_{hf}, W_{hi}, and W_{ho} are the weight matrices of the previous time step's hidden states, while W_{xg}, W_{xf}, W_{xi}, and W_{xo} are the weight matrices of the current time step's input. The superscript 's' represents each type of features in the input sequences, the subscript 't' represents the time step, σ represents the sigmoid function, the operator '$*$' indicates the matrix multiplication, and the operator '$\odot$' indicates elementwise multiplication.

Residual learning, which was first introduced in image recognition for training ultra-deep CNNs [30], is adopted in this model for the representation learning of higher layers and the reformulation of the layers by learning residual functions with reference to the layer inputs. The formula of residual learning can be expressed as follows:

$$y = F(x, W) + x,$$

where x and y refer to the input and output vectors of the layers under consideration in the multimodal residual LSTM network, respectively, and $F(x, W)$ represents the residual function learned by the corresponding layers.

Through residual learning, the output of the corresponding layer becomes a linear combination of the input and a nonlinear residual. Residual learning provides a shortcut across the layers for training the multilayer LSTM network more effectively and avoids the problem of vanishing gradients due to the multilayer structure by adjusting the residual $F(x, W)$.

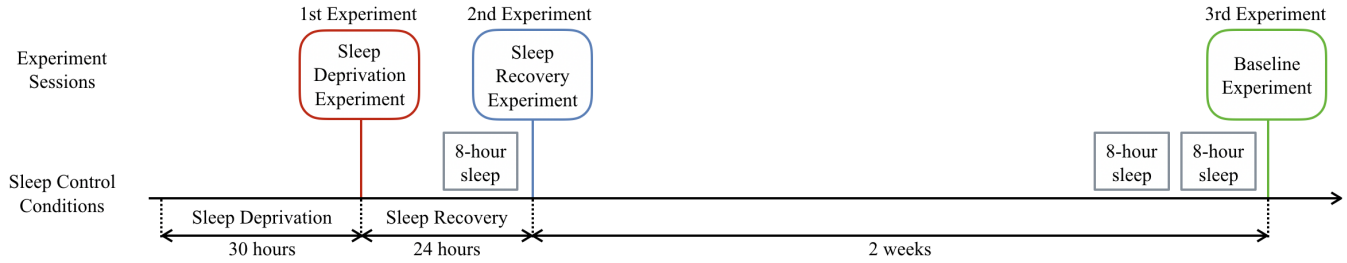


Fig. 2. Illustration of our proposed experiment design. The experiment consists of a sequence of experiment sessions of three different sleep control conditions: sleep deprivation condition, sleep recovery condition and baseline condition. In the first session, a sleep deprivation experiment is conducted after 30-hour sleep deprivation. In the second session, sleep recovery experiment is then conducted after an 8-hour sleep recovery. In the third session, a baseline experiment is carried out at least 14 days after the second session. During this 14-day period, the subjects sleep according to their sleep routines and, in particular, the 8-hour sleep condition is required during the 2 nights prior to the third experiment session.

In each layer of the multimodal residual LSTM Network, layer normalization [31] is applied to stabilize the hidden-state dynamics and reduce the training time of deep RNNs by recentering and rescaling the neurons of the LSTM as follows:

$$\mu_t = \frac{1}{H}\sum_{i=1}^{H}(h_t)_i,$$

$$\delta_t = \sqrt{\frac{1}{H}\sum_{i=1}^{H}((h_t)_i - \mu_t)^2},$$

$$y_t = f(\frac{g}{\delta_t} \odot (h_t - \mu_t) + b),$$

where $(h_t)_i$ represents the hidden state of the ith neuron in each LSTM layer, the subscript 't' represents the time steps, and g and b are trainable weights with the same size as h_t that are used for rescaling and recentering the input of the activation function f, respectively.

Dropout in each layer of the multimodal residual LSTM network is applied before the forward connections to reduce overfitting. High-level representations of three types of features are eventually concatenated to predict the emotion labels by dense layer with Softmax activation.

This multimodal residual LSTM network can effectively capture the intramodality correlations by each 4-layer LSTM structure and the intermodality correlations among the three types of features by sharing weights across the three LSTM structures. Residual learning and layer normalization are also employed for efficient training.

IV. Experiments

The sleep deprivation experiments consist of three experiment sessions with different sleep conditions, namely, the sleep deprivation experiment, the sleep recovery experiment and the baseline experiment, corresponding to the sleep conditions depicted in Fig. 2.

Sixteen healthy subjects (eight males and eight females, age range: 18-32 years, mean: 22.25, std: 3.09) participated in the experiments. All of the subjects were preselected to ensure that they had regular daily sleep routines and had the habit of sleeping for 7-8 hours every day. The subjects satisfying these conditions are considered more obviously influenced by sleep deprivation compared to people regularly stay up late. Prior to each experiment session, the subjects were informed of the experimental purpose, the experimental procedure, and the harmlessness of the equipment used in the experiment. The study was approved by the local ethics committee.

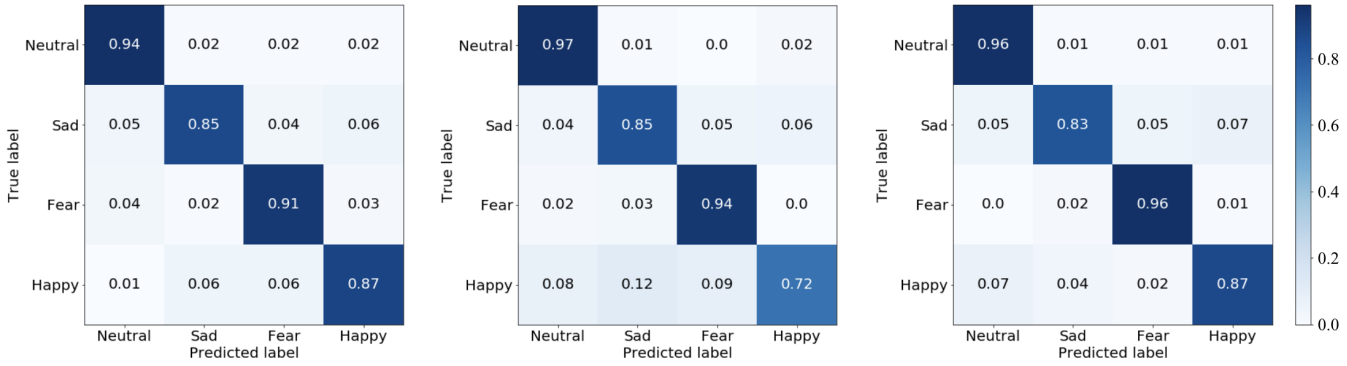
In the first experiment session, the sleep deprivation experiment was conducted after 30-hour sleep deprivation. After one normal night sleep recovery of 8-hour sleep, the subjects participated in the second experiment session, namely, the sleep recovery experiment. The third experiment session was the baseline experiment. To thoroughly eliminate the influence of previous sleep deprivation, the baseline experiment was conducted at least 14 days after the second session. The subjects were required to maintain their regular sleep routines during the 14 days between the second session and the third session and to sleep for 8 hours during the 2 nights before the baseline experiment. The sleep conditions of all of the experiment sessions were monitored and recorded by portable smart bands that tracked sleep duration and sleep quality information of the subjects.

The emotion recognition task of each experiment session is watching emotion stimuli film clips. The emotion stimuli film clips used in our experiments are exactly same as the stimuli clips used in a public emotion EEG dataset called the SEED-IV Dataset [21]. All of the emotion stimuli film clips contain highly emotional contents and have been demonstrated to be reliable for eliciting the target emotions. Each experiment session contains 24 different trials (six trials per emotion) of stimuli clips that are designed to elicit four target emotions: happiness, fear, sadness and neutral. Each stimuli clip was presented only once during the three experiment sessions to avoid repetition.

EEG signals were recorded by a 62-channel wet electrode cap at a sampling rate of 1000 Hz using the ESI NeuroScan system. The electrodes on the cap are placed according to the higher-resolution international 10-20 system. Eye movement signals were simultaneously collected by SMI-ETG eye-tracking glasses.

TABLE II
ACCURACY (%) OF SUBJECT-DEPENDENT EMOTION CLASSIFICATION TASKS

Experiment	Model	Feature	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Mean ± Std
Baseline	SVM	EEG DE	53.12	57.76	57.24	68.77	65.40	51.71	59.00 ± 13.18
	SVM	EEG Strength	53.02	61.01	54.5	67.14	56.51	40.40	55.43 ± 11.13
	SVM	EEG DE, Eye	47.32	52.98	71.48	75.47	64.71	46.97	59.83 ± 14.40
	Residual LSTM	EEG DE	86.31	84.62	87.95	88.47	89.71	79.74	86.13 ± **3.54**
	Multimodal LSTM	EEG DE, Eye	87.10	83.78	93.75	96.75	87.20	73.66	87.04 ± 6.18
	Multimodal LSTM	EEG DE, EEG Strength, Eye	89.96	86.38	93.75	95.98	92.86	76.22	**89.19** ± 6.29
Sleep Deprivation	SVM	EEG DE	50.22	49.45	42.29	59.23	51.59	54.71	51.25 ± 12.78
	SVM	EEG Strength	43.70	54.14	46.83	55.65	57.66	55.68	52.28 ± 13.95
	SVM	EEG DE, Eye	46.75	40.58	49.88	54.64	49.93	44.59	47.73 ± 8.86
	Residual LSTM	EEG DE	81.62	79.53	80.38	88.07	88.47	87.30	84.23 ± 7.93
	Multimodal LSTM	EEG DE, Eye	84.82	89.81	78.89	85.81	85.66	88.29	85.55 ± **5.02**
	Multimodal LSTM	EEG DE, EEG Strength, Eye	89.73	89.73	80.06	87.85	86.73	87.08	**86.86** ± 5.55
Sleep Recovery	SVM	EEG DE	55.73	44.62	40.38	61.04	53.37	61.16	52.72 ± 13.77
	SVM	EEG Strength	42.76	51.54	48.41	51.76	41.96	48.34	47.46 ± 11.77
	SVM	EEG DE, Eye	50.16	64.18	53.97	67.46	65.00	64.81	60.93 ± 12.07
	Residual LSTM	EEG DE	86.58	77.45	80.01	88.37	84.65	88.49	84.26 ± **6.37**
	Multimodal LSTM	EEG DE, Eye	86.66	88.07	80.08	95.61	92.46	88.34	88.54 ± 7.36
	Multimodal LSTM	EEG DE, EEG Strength, Eye	87.04	90.82	83.53	97.54	94.42	90.60	**90.66** ± 6.95



(a) (b) (c)

Fig. 3. Confusion matrices of multimodal residual LSTM networks using EEG DE features, EEG strength features with correlation connectivity and eye movement features. (a) Baseline experiment, (b) sleep deprivation experiment, and (c) sleep recovery experiment.

V. Results And Discussions

A. Subject-dependent Classification Performance

We first evaluate the performance of our model on the subject-dependent emotion classification tasks. For subject-dependent emotion classification tasks, the models were trained and tested individually for each session of a single subject. Six-fold cross-validation was applied and each fold contains four emotional stimuli trials corresponding to the four target emotions. The average classification accuracy and standard deviation over subjects and folds were calculated. A linear kernel support vector machine (SVM) with default parameters implemented by LIBLINEAR [33] with EEG DE features, EEG strength features, and the concatenation of EEG DE features and eye movement features as inputs was adopted as the baseline classifier. For the multimodal residual LSTM networks, we adopted two input cases. In the first case, we used the EEG DE features and eye movement features as the inputs of two parallel weight-sharing LSTM structures. In the second case, the strength features with correlation connectivity were applied as the third type input features of the additional parallel multimodal residual LSTM structure. We determined the optimal parameter settings for the multimodal residual LSTM network. The LSTM node number was set to 128, the layer number to 4, dropout ratio to 0.5, l2 regularization to 1e-2, learning rate to 1e-3, and the maximum number of epochs to 1200. The Adam optimization algorithm was used to train the network. The accuracy not increasing by 0.1% on the validation set for the previous 30 epochs was used as the early stop criterion. The subject-dependent classification accuracies and standard deviations of three experiment sessions of the respective sleeping conditions are summarized in Table II and Fig. 3.

Both the residual LSTM network using only the EEG DE features as inputs and the multimodal residual LSTM networks achieve clearly superior emotion classification performance compared to that of the baseline SVM model. Using the EEG DE features, EEG strength features and eye movement features as inputs, the multimodal residual LSTM network obtained the

best subject-dependent classification performance characteristics for all three sessions, with the mean accuracies of 89.19%, 86.86%, and 90.66%, respectively, which is 2.15%, 1.31%, and 2.12% higher than the multimodal residual LSTM network using the EEG DE and eye movement as input features. The standard deviations of the multimodal residual LSTM network are also observed to be obviously lower compared to SVM for all three sessions, suggesting that the multimodal residual LSTM network is more stable than the baseline SVM model.

As shown in Table II, the mean accuracy of the classification performance using the best multimodal residual LSTM networks of the sleep deprivation session is 2.33% and 3.80% lower compared to the other two experiment sessions. The main reason for this accuracy reduction is the classification performance of the happiness emotion that is depicted by the confusion matrices of the three experiment sessions in Fig. 3. It is observed that the mean classification accuracy of the happiness emotion for the sleep deprivation session is 72%, dropping by 15% compared to the baseline and the sleep recovery sessions. This indicates that sleep deprivation impairs the stimulation of the happiness emotion, and one night of sleep recovery can reactivate the stimulation of the happiness emotion, which is consistent with a previous study [34]. Moreover, among the three experiment sessions, the fear emotion and the neutral emotion can all be identified with relatively high accuracy. This observation indicates that the emotion patterns of these two emotions are insensitive to sleep deprivation. The sadness emotion is also stable across the different sleep experiment sessions but is more difficult to be distinguished than the fear emotion and the neutral emotion. Furthermore, the accuracy distributions of the confusion matrices of baseline and sleep recovery sessions are quite similar, so we can infer that the overall emotional states of the subjects are generally recovered to the baseline standard from the preceding sleep deprivation through sleep recovery.

B. Cross-subject Classification Performance

The results of the cross-subject emotion classification performance are presented in Table III. Leave-one-subject-out cross-validation (*i.e.*, 16-fold validation) was applied to compare the classification performance of the multimodal residual LSTM network using different input features. It is observed that for all three experiment sessions, using the combination of DE, strength and eye movement features improves the classification performance compared to using DE and eye movement features only. These results demonstrate the effectiveness of strength features across the individual differences. Moreover, for all three sessions, the reduction in accuracy and the increase in the standard deviation compared to the subject-dependent classification tasks indicate the individual differences on emotion patterns, whereas the cross-subject classification performance of sleep deprivation session is highest among the three sessions and decreases 4.83% compared to the subject-dependent task, which is much less than the performance reduction of the other two sessions. We can infer that the emotion patterns under sleep deprivation have certain general characteristics, regardless of individual differences.

TABLE III
ACCURACY (%) OF CROSS-SUBJECT EMOTION CLASSIFICATION TASKS USING MULTIMODAL RESIDUAL LSTM NETWORK

Experiment	Feature	Mean ± Std
Baseline	EEG DE, Eye	76.38 ± 9.64
	EEG DE, EEG Strength, Eye	**77.67 ± 8.67**
Sleep Deprivation	EEG DE, Eye	81.04 ± 8.24
	EEG DE, EEG Strength, Eye	**82.03 ± 8.24**
Sleep Recovery	EEG DE, Eye	80.93 ± **9.75**
	EEG DE, EEG Strength, Eye	**81.99** ± 10.25

C. Topographic Neural Patterns

The average energy distributions for the happiness, sadness, fear, and neutral emotions in the gamma band of the EEG DE features are depicted in Fig. 4 because the most distinguishable neural patterns are observed in this band. Under sleep deprivation, for the happiness emotion and sadness emotion, the prefrontal area is least activated compared to the baseline condition and the sleep recovery condition; for neutral emotion, energy is concentrated in the lateral temporal area, which is most likely due to the lack of attention caused by sleep deprivation.

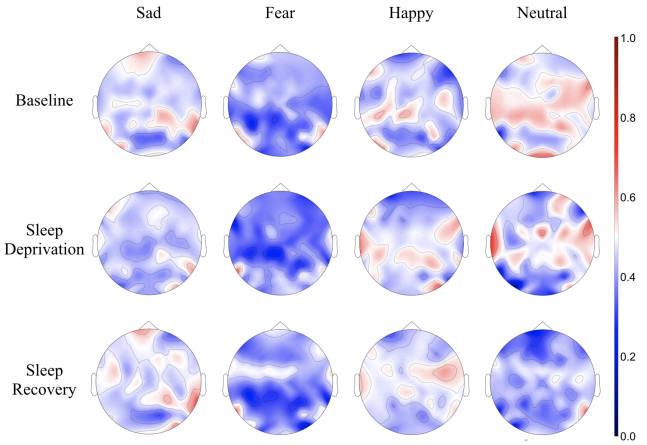


Fig. 4. Topographic maps of the four emotions (sadness, fear, happiness, and neutral) in the gamma band for the three sessions.

For all the three experiment sessions, the activation levels of the prefrontal area for the happiness emotion are lower than those for the sadness emotion. Fear emotion is generally low-activated except for the post lateral temporal area, and this characteristic is stable for all three sleep conditions.

VI. Conclusions

In this paper, we have introduced the multimodal residual LSTM network to investigating emotion recognition under sleep deprivation with the integration of the EEG DE features, EEG strength features with correlation connectivity, and eye movement features. For each type of feature, four LSTM layers

are employed for explicitly learning the high-level temporal features. The weight-sharing architecture in each layer across the parallel LSTM structures corresponding to three type of features reinforces the effectiveness of learning intramodality and intermodality correlations. The experimental results on the sleep deprivation session, the sleep recovery session and the baseline session demonstrate that the classification accuracy increases with increasing number of the modalities used for both subject-dependent and cross-subject emotion recognition tasks under sleep deprivation. The elicitation of the happiness emotion is the most impaired by sleep deprivation compared with the other emotion types. Moreover, under sleep deprivation, the prefrontal brain area is less activated for the happiness emotion and sadness emotion in the gamma band, whereas fear emotion corresponds to a highly robust neural pattern.

ACKNOWLEDGMENTS

This work was supported in part by the National Key Research and Development Program of China (2017YFB1002501), the National Natural Science Foundation of China (61673266 and 61976135), SJTU Trans-med Awards Research (WF540162605), the Fundamental Research Funds for the Central Universities, and the 111 Project.

REFERENCES

[1] D. Zohar, O. Tzischinsky, R. Epstein, and P. Lavie, "The effects of sleep loss on medical residents' emotional reactions to work events: a cognitive-energy model," Sleep, 28(1), pp. 47-54, 2005.

[2] M.P. Walker and E. Van Der Helm, "Overnight therapy? The role of sleep in emotional brain processing," Psychological Bulletin, 135(5), pp. 731, 2009.

[3] A.N. Goldstein-Piekarski, S.M. Greer, J.M. Saletin, and M.P. Walker, "Sleep deprivation impairs the human central and peripheral nervous system discrimination of social threat," Journal of Neuroscience, 35(28), pp. 10135-10145, 2015.

[4] D.F. Dinges, F. Pack, K. Williams, K.A. Gillen, J.W. Powell, G.E. Ott, C. Aptowicz, and A.I. Pack, "Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4–5 hours per night," Sleep, 20(4), pp. 267-277, 1997.

[5] S. Pallesen, B.H. Johnsen, A. Hansen, J. Eid, J.F. Thayer, T. Olsen, and K. Hugdahl, "Sleep deprivation and hemispheric asymmetry for facial recognition reaction time and accuracy," Perceptual and Motor Skills, 98(3 suppl), pp. 1305-1314, 2004.

[6] U. Wagner, N. Kashyap, S. Diekelmann, and J. Born, "The impact of post-learning sleep vs. wakefulness on recognition memory for faces with different facial expressions," Neurobiology of Learning and Memory, 87(4), pp. 679-687, 2007.

[7] S.S. Yoo, N. Gujar, P. Hu, F.A. Jolesz, and M.P. Walker, "The human emotional brain without sleep—a prefrontal amygdala disconnect," Current Biology, 17(20), pp. R877-R878, 2007.

[8] P.L. Franzen, D.J. Buysse, R.E. Dahl, W. Thompson, and G.J. Siegle, "Sleep deprivation alters pupillary reactivity to emotional stimuli in healthy young adults," Biological Psychology, 80(3), pp. 300-305, 2009.

[9] P. Tzirakis, G. Trigeorgis, M.A. Nicolaou, B.W. Schuller, and S. Zafeiriou, "End-to-end multimodal emotion recognition using deep neural networks," IEEE J. Sel. Topics Signal Process., vol. 11, no. 8, pp. 1301–1309, Dec. 2017.

[10] W.-L. Zheng, B.-N. Dong, and B.-L. Lu, "Multimodal emotion recognition using EEG and eye tracking data," In Proc. Int. Conf. IEEE Eng. Med. Biol. Soc., pp. 5040–5043, 2014.

[11] Y. Lu, W.-L. Zheng, B. Li, and B.-L. Lu, "Combining eye movements and EEG to enhance emotion recognition," in Proc. Int. Joint Conf. Artif. Intell., pp. 1170–1176, 2015.

[12] J.-X Ma, H. Tang, W.-L Zheng, and B.-L. Lu, "Emotion Recognition using Multimodal Residual LSTM Network," In Proceedings of the 27th ACM International Conference on Multimedia, pp. 176-183, 2019.

[13] J. Ren, Y. Hu, Y.W. Tai, C. Wang, L. Xu, W. Sun, and Q. Yan, "Look, listen and learn—a multimodal LSTM for speaker identification," In Thirtieth AAAI Conference on Artificial Intelligence, March 2016.

[14] L.-C. Shi, Y.-Y. Jiao, and B.-L. Lu, "Differential entropy feature for EEG-based vigilance estimation," In Proc. Int. Conf. IEEE Eng. Med. Biol. Soc., pp. 6627–6630, 2013.

[15] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," IEEE Transactions on Autonomous Mental Development, vol. 7, no. 3, pp. 162–175, 2015.

[16] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," IEEE Transactions on Affective Computing, pp. 99, 1-1, 2018.

[17] M. Chen, J. Han, L. Guo, J. Wang, and I. Patras, "Identifying valence and arousal levels via connectivity between EEG channels," International Conference on Affective Computing and Intelligent Interaction (ACII), pp. 63–69, 2015.

[18] Y.-Y. Lee and S. Hsieh, "Classifying different emotional states by means of EEG-based functional connectivity patterns," PLoS One, vol. 9, no. 4, p. e95415, 2014.

[19] X. Wu, W.-L. Zheng, and B.-L. Lu, "Identifying Functional Brain Connectivity Patterns for EEG-Based Emotion Recognition," International IEEE/EMBS Conference on Neural Engineering (NER), San Francisco, CA, USA, pp. 235-238, 2019.

[20] N. Mahendran, P.-D.-R. Vincent, K. Srinivasan, V. Sharma, and D.K. Jayakody, "Realizing a Stacking Generalization Model to Improve the Prediction Accuracy of Major Depressive Disorder in Adults," IEEE Access, 8, pp. 49509-49522, 2020.

[21] W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. Cichocki, "Emotionmeter: A multimodal framework for recognizing human emotions," IEEE Transactions on Cybernetics, vol. 49, pp. 1110–1122, 2019.

[22] W. Lin, C. Li, and S. Sun, "Deep convolutional neural network for emotion recognition using EEG and peripheral physiological signal," In International Conference on Image and Graphics, pp. 385-394, Springer, Cham, September 2017.

[23] X. Zhang, J. Pan, J. Shen, Z.-U. Din, J. Li, D. Lu, M. Wu, and B. Hu, "Fusing of Electroencephalogram and Eye Movement with Group Sparse Canonical Correlation Analysis for Anxiety Detection," IEEE Transactions on Affective Computing, 2020.

[24] L.-C. Shi and B.-L. Lu, "Off-line and on-line vigilance estimation based on linear dynamical system and manifold learning," In Proc. Int. Conf. IEEE Eng. Med. Biol. Soc., pp. 6587–6590, 2010.

[25] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: uses and interpretations," Neuroimage, vol. 52, no. 3, pp. 1059–1069, 2010.

[26] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 8, pp. 1226–1238, 2005.

[27] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, 9(8), pp. 1735-1780, 1997.

[28] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: a model for automatic sleep stage scoring based on raw single-channel EEG," IEEE Transactions on Neural Systems and Rehabilitation Engineering, 25(11), pp. 1998-2008, 2017.

[29] H. Tang, W. Liu, W.-L. Zheng, and B.-L. Lu, "Multimodal emotion recognition using deep neural networks," In International Conference on Neural Information Processing, pp. 811-819, Springer, Cham, 2017.

[30] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770-778, 2016.

[31] J.L. Ba, J.R. Kiros, and G.E. Hinton, "Layer normalization," arXiv preprint arXiv:1607.06450, 2016.

[32] R. Jadhav, V. Chellwani, S. Deshmukh, and H. Sachdev, "Mental Disorder Detection: Bipolar Disorder Scrutinization Using Machine Learning," In 9th International Conference on Cloud Computing, Data Science Engineering (Confluence), pp. 304-308, IEEE, 2019.

[33] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: A library for large linear classification," Journal of Machine Learning Research, vol. 9, no. Aug, pp. 1871–1874, 2008.

[34] E. Van Der Helm, N. Gujar, and M.P. Walker, "Sleep deprivation impairs the accurate recognition of human emotions," Sleep, 33(3), pp. 335-342, 2010.