University of Reading

Emotion classification on eye-tracking and electroencephalograph fused signals employing deep gradient neural networks

Article

Accepted Version

Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Wu, Q., Dey, N., Shi, F., González Crespo, R. and Sherratt, S. ORCID: https://orcid.org/0000-0001-7899-4445 (2021) Emotion classification on eye-tracking and electroencephalograph fused signals employing deep gradient neural networks. Applied Soft Computing, 110. 107752. ISSN 1568-4946 doi: 10.1016/j.asoc.2021.107752 Available at https://centaur.reading.ac.uk/99520/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.1016/j.asoc.2021.107752

Publisher: Elsevier

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

University of Reading

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Emotion classification on eye-tracking and electroencephalograph fused signals employing deep gradient neural networks

Qun Wu[1, *], Nilanjan Dey[2], Fuqian Shi[3], Rubén González Crespo[4], and R. Simon Sherratt[5]

1. Institute of Universal Design, Zhejiang Sci-Tech University, Hangzhou, PR China
2. Department of Computer Science and Engineering, JIS University, Kolkata, 700109. India
3. Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, 08903, USA
4. Department of Computer Science and Technology, Universidad Internacional de La Rioja, Logroño, Spain
5. Department of Biomedical Engineering, the University of Reading, RG6 6AY, U.K.

***Corresponding authors:**
Qun Wu, Address: No. 928, the 2nd Street, Xiasha High-Tech Education Zone, Hangzhou City, Zhejiang, PR China; E-mail: wuq@zstu.edu.cn; Tel.: +86-571-86843291

Abstract
Emotion produces complex neural processes and physiological changes under appropriate event stimulation. Physiological signals have the advantage of better reflecting a person's actual emotional state than facial expressions or voice signals. An electroencephalogram (EEG) is a signal obtained by collecting, amplifying, and recording the human brain's weak bioelectric signals on the scalp. The eye-tracking (E.T.) signal records the potential difference between the retina and the cornea and the potential generated by the eye movement muscle. Furthermore, the different modalities of physiological signals will contain various information representations of human emotions. Finding this different modal information is of great help to get higher recognition accuracy. The E.T. and EEG signals are synchronized and fused in this research, and an effective deep learning (DL) method was used to combine different modalities. This article proposes a technique based on a fusion model of the Gaussian mixed model (GMM) with the Butterworth and Chebyshev signal filter. Features extraction on EEG and E.T. are subsequently calculated. Secondly, the self-similarity (SSIM), energy (E), complexity (C), high order crossing (HOC), and power spectral density (PSD) for EGG, and electrooculography power density estimation ((EOG-PDE), center gravity frequency (CGF), frequency variance (F.V.), root mean square frequency (RMSF) for E.T. are selected hereafter; the max-min method is applied for vector normalization. Finally, a deep gradient neural network (DGNN) for EEG and E.T. multimodal signal classification is proposed. The proposed neural network predicted the emotions under the eight emotions event stimuli experiment with 88.10% accuracy. For the evaluation indices of accuracy (Ac), precision (Pr), recall (Re), F-measurement (Fm), precision-recall (P.R.) curve, true-positive rate (TPR) of receiver operating characteristic curve (ROC), the area under the curve (AUC), true-accept rate (TAR), and interaction on union (IoU), the proposed method also performs with high efficiency compared with several typical neural networks including the artificial neural network (ANN), SqueezeNet, GoogleNet, ResNet-50, DarkNet-53, ResNet-18, Inception-ResNet, Inception-v3, and ResNet-101.

Keywords: fused deep neural network, electroencephalogram, eye-tracking, signal process, Gaussian mixed model, emotion stimuli

1. Introduction

As the most easily obtained signals of the human body through sensors, physiological signals contain many important physiological and psychological information. Getting and identifying a variety of physiological signals is of great significance for computers to recognize human emotions. Limited by the complexity of physiological signals such as brain electricity, skin galvanic, respiration, pulse, etc., it has always been challenging to extract useful features from these physiological signals and accurately classify and predict emotions. With the introduction of the concept of "affective computing", scholars have devoted themselves to the mathematicalization of the idea of emotion so that computers can recognize and process and recognize and classify emotional states. Physiological signals, including brain electricity, electrocardiogram, skin electricity, etc., are crucial input signals in affective computing. Emotional computing is an interdisciplinary research field involving multiple disciplines such as computer science, psychology, and cognitive science. It aims to research and develop the ability to recognize, interpret, process, and simulate human emotions [1]. With the continuous development of human-computer interaction, machine learning, deep learning, etc., emotional computing is used in healthcare, media entertainment, information retrieval, education, and intelligent wearable devices. All have broad application prospects.

Russell established a two-dimensional emotion model and emotion ring theory based on arousal degree and pleasure degree so that computers can recognize the concept of emotion [2]. Conte and Plutchik divided emotions into eight basic categories: anger, fear, sadness, disgust, expectation, surprise, approval, and happiness [3]. The discrete emotion classification method is relatively simple and easy to understand and has been widely used. Changes in physiological signals usually accompany human emotional changes. Physiological signals are controlled by the autonomic nervous and endocrine system, not controlled by subjective consciousness, and may objectively and truly reflect individual physiological, mental, and emotional states [4,5]. Physiological signals used in emotion computing mainly include electroencephalogram (EEG), eye tracking (E.T.), Electromyography (EMG), electrocardiogram (ECG), and respiratory signals [6,7]. However, recent research is indicating the advantages of using cortisol biosensors [8]

Eye movement research is to record the movement track of the user's eye line when performing a specific task with the help of an eye movement instrument to analyze the user's cognitive behavior and psychological activities. Eye-tracking is to measure the fixation point (the place of fixation) or the movement of the eye. An eye tracker is used in visual system research, psychology, linguistics, marketing, human-computer interaction input devices, and product design. An eye tracker is also increasingly used in rehabilitation and assistive applications (e.g., related to the control of wheelchairs, robotic arms, and prosthetics). The image processing technology is generally used to locate the pupil position, obtain coordinates, and calculate the eye gaze or gaze point through a particular algorithm. At present, popular eye-tracking technology is mainly based on the "non-invasive" technology of eye video analysis (VOG) [9]. Its basic principle aims at a beam of light (near-infrared light) and a camera at the subject's eyes, infer the subject's gaze direction through light and back-end analysis, and the camera records the interaction process.
For this reason, current researchers usually use eye trackers based on the video, such as Eyelink 1000 plus. A video-based eye-tracker can monitor gaze and display other useful measurement indexes, including pupil size and blink rate. Emotion recognition is currently widely used in neuromarketing, entertainment, health, and psychology. Recently, the development trend of the system is to improve the ability of emotion detection and response. Because emotion contains many nonverbal cues, and previous studies used different ways as indicators of emotional state. Emotion detection has developed in many safe-driving, mental health monitoring, and social security [10].

2 Literature Reviews

Many types of research focus on the interaction between users and computers. Therefore, human-computer interaction (HCI) has become more and more critical in computer science. HCI plays an essential role in identifying, detecting, processing, and responding to users' emotions [11][12]. Fischer et al. [13] and Cowie et al. [14] focus on user modeling and emotion recognition in HCI. The computer system that can detect human emotions is called an emotional computer system [15]. Affective computing is a research field that combines computer science, psychology, cognitive science, artificial intelligence, and devices that can recognize, read, store, and respond to human emotions. It has become an important research area to develop a system that can automatically recognize human emotions. This study is a review of affective computing [16]. The emotion recognition system also applied E.T., EEG, and galvanic skin response (GSR) combination data recently [17, 18]. Eye-tracking is a process of measuring when and where the user's eyes focus, in other words, the gaze point and pupil size. An eye tracker is used to measure the eye position and eye movement of individuals. It is a sensor technology that can better understand the user's visual attention. Monitor the light source (for example, the pupil reflection characteristics of the eye). Hess et al. proposed an increase in pupil size, which is accompanied by emotional arousal or fascinating visual stimulation; besides pupil diameter, many other functions can be used to recognize emotions, including gaze time, saccade, and EOG signals [19]. Because eye movement signals can be used as user behavior indicators, they are widely used in HCI research [20]. Most previous studies used eye movement to analyze user interest, visual search process, and information processing. As a result, eye tracking and emotion processing are becoming more and more popular in cognitive science. There are also reports on previous work using eye movement analysis to use EOG signals for activity recognition rather than emotion recognition; eye movement signals enable us to accurately identify the reasons that attract users' attention and observe their subconscious behavior [21]. It may be an important clue in a context-aware environment, including supplementary information for emotion recognition. The signal can provide some emotion-related features to determine the user's emotional state. Researchers can estimate the user's mood according to the change of pupil size.

Researchers can gain more cognitive understanding by analyzing eye movement data, such as human fixation position and fixation time, by integrating eye tracking with other neuroimaging techniques such as EEG, fNIRS, and fMRI [22]

[23]. Because eye movement data reflect the behavior, it cannot directly reflect the cognitive and thinking process. Therefore, eye movement research needs to cooperate with good experimental design to correctly interpret eye movement data or cooperate with interviews or backtracking tests to reflect the cognitive and thinking process. EEG is a spontaneous and rhythmic activity of the nerve. Its frequency range is 1-30 times per second. It can be divided into four bands: δ (1-3hz), θ (4-7hz), α (8-13hz), and β (14-30hz). Besides, when awakening and focusing on something, we often see a gamma wave with a higher frequency than the beta wave, whose frequency is 30 -80Hz, and the amplitude range is variable; while in sleep, there are other regular brain waves with particular waveforms, such as hump wave, pulse wave, kappa - complex wave, μ wave, etc. The available features include the time domain, frequency domain, time-frequency domain, and spatial domain. The standard database of emotional stimulation includes SEED, DREAMER AND CAS-THU [24, 25]. The characteristics of the calculation include the time domain features (mean, standard deviation, first-order difference, normalized first-order difference, second-order difference, normalized second-order difference, Hjorth characteristic (activity, mobility, complexity), energy, power, high-order zero-crossing analysis, instability index, Higuchi fractal dimension), time-frequency domain (power spectral density, higher-order spectrum, differential entropy), and spatial domain (differential asymmetry (DASM), rational asymmetry (RASM), asymmetry coefficient, differential causality (DCAU), multi-dimensional directed information) [26, 27].

The research framework includes establishing a standardized library of emotions and the way of audio-visual stimulation to induce emotions. Moreover, recently, some researchers have focused on different combinations of physiological signals on emotional computing. Many researchers use the deep learning method to study the fusion of multimodal physiological signals to improve the impact of emotional computing tasks [28, 29]. Different modal physiological signals contain different information representations of human emotion. Finding the correlation between this different modal information is of great help to improve emotion recognition accuracy. A practical method is to use deep learning to fuse different modal physiological signals [30]. In general, the multimodal framework of emotion recognition based on EEG electrodes and eye movement signals can be easily embedded into wearable devices such as earphones or eyeglasses frames. Four emotion categories, happiness, sadness, fear, and routine, are induced and classified by movie clips. The principal component analysis is used to preprocess the eye movement signals; the features extract five kinds of eye movement parameters from the eye movement signals. Then the feature level fusion of EEG features and eye movement features is used as the benchmark of modal fusion; that is, the EEG features and eye movement features are directly connected into a more extensive feature vector [31].

Researchers commonly used the deep neural network model in emotional computing tasks based on physiological signals for the fused signals. A deep neural network is an external neural network with multiple hidden layers [32]. Each neurone in the hidden layer is connected to many other neurons. Each arrow is attached with a weight attribute, which controls the extent to which neurons' activation affects other neurons connected with it. The deep learning of these words comes down to the deep learning effect [33]. The number of hidden layers selected depends on the nature of the problem and the dataset's size [34]. A deep neural network (DNN) is a discriminant model, which a backpropagation algorithm can train. DNN is the basis of many A.I. applications. Due to the breakthrough application of DNN in speech recognition and image recognition, the application of DNN has explosive growth. DNN is deployed in applications ranging from autopilot, cancer detection to complex games. In many of these areas, DNN can surpass human accuracy. This is different from previous methods using manual feature extraction or expert design rules [35].

Generally, the objective function of a deep neural network is a non-convex function, which means that the global minimum of the role is difficult to find. The current optimization methods can only see the local minimum of the objective function. Due to its multiple nonlinear transformations, neural networks are difficult to use analytical techniques to solve all at once. The stochastic gradient descent algorithm can be said to be the cornerstone of deep learning. This seemingly simple optimization method supports the entire deep learning and various applications of deep neural networks. As long as it involves deep neural network training, it is used for gradient descent. Stochastic gradient descent uses an iterative method to make complex neural network training feasible, so an improvement of the gradient should be developed in this research. Physiological signal classification by using deep neural network in this paper established a data analysis model for one-dimensional physiological signal feature extraction and state recognition based on deep learning, "pre-training + fine-tuning" was adopted in the training process; the pre-training process adopts bottom-up unsupervised training, and first trains the first hidden signal. Then the next hidden layer is trained layer by layer, and the output of the previously hidden layer node is taken as the input; the output of this hidden layer node is taken as the input of the next hidden layer; In the pre-training stage, the first weight is trained, and then the trained node is used as the input of the second, and so on. After all, training is completed, the backpropagation

(B.P.) algorithm is used to fine-tune the network. Finally, the output feature vectors of the deep belief network are input into a "softmax" classifier to judge the individual state of the one-dimensional physiological signal. The framework of the work is shown in Figure 1.

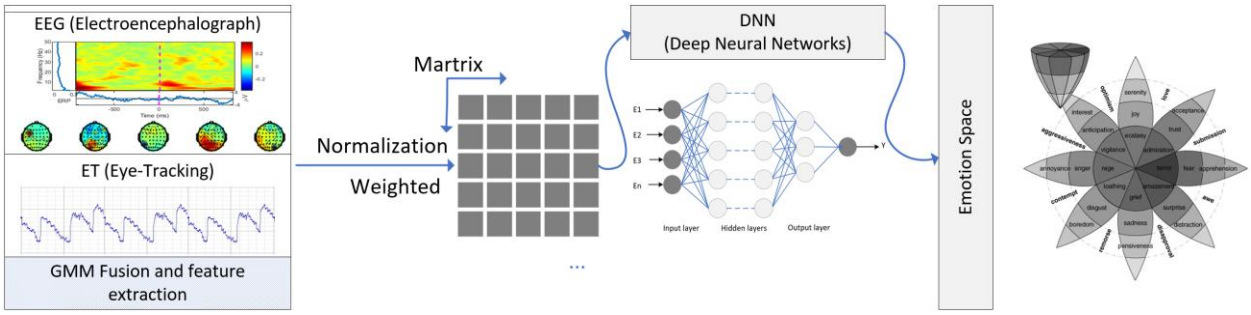


Figure 1. the framework of affective computing by using eye-tracking and electroencephalograph fused signals.

3. Modeling

3.1 Fused signal processing

To clean the signal, ordinary filters include B filter (Butterworth), C filter (Chebyshev), and Bessel filter [36] were used in this research. To filter the eye movement signal, B low-pass filter is used subsequently. Supposed that the signal is denoted by $x(t)$, the transformation on signal is defined by,

$$|H(\varpi)|=\frac{1}{\sqrt{1+\varepsilon^2(\frac{\varpi}{\varpi_p})^{2n}}} \tag{1}$$

Where $_n$ is the order of the filter ϖ_p is bandwidth frequency? For multisource signals, the fusion method here is a Gaussian Mixed Model, a multivariate normal distribution. For signal fusion task, the basic model for single Gaussian is that,

$$f(x(t);\sigma,\mu)=\frac{1}{\sqrt{(2\pi)^n|\sigma|}}e^{[-\frac{1}{2}(x(t)-\mu)^T\sigma^{-1}(x(t)-\mu)]} \tag{2}$$

$x(t)$ A signal μ is an expectation on $x(t)$, σ is a covariance deviation, n is the order of signal (dimensions).

Now we deployed the multisource signal fused method by using signal GMM. The first step is to set Q function as,

$$Q(\theta,\theta^t)=\sum_Y[\log P(X,Y|\theta)|X,\theta^{(t)}] \tag{3}$$

Transferring missing signals dataset by using,

$$P(Y_j\in M_k|X_j,\Theta^{(t)})=\frac{w_k^{(t)}f(X_j|\theta_k^{(t)})}{\sum_{k=1}^{K}w_k^{(i)}f(X_j|\theta_k^{(t)})} \tag{4}$$

Where $Y_j\in M_k$ presents the j-th observation on the k-th model of signal. Write out the criterion function of the mixed distribution model as,

$$Q(\Theta,\Theta^{(t)})=\sum_{j=1}^{N}\sum_{k=1}^{K}\log(w_k)P(Y_j,\Theta^{(t)})+\sum_{j=1}^{N}\sum_{k=1}^{K}\log(f_k(X_j|Y_j\in M_k,\theta_k))P(Y_j\in M_k|X_j,\Theta^{(t)}) \tag{5}$$

Where, $\theta_k=[\mu_k,\sigma_k]$ are the parameters of the distribution k. $\Theta=\{\theta_1,\theta_2,...,\theta_K\}$ is a parameter set. N is a number of sample signals. K is several models. To estimate the parameters in the mixed model in Eq. (5), we first optimize the weights w_k subject to $\sum_k w_k=1$. By using Lagrange, multiple operators, we have that,

$$J_w = \sum_{j=1}^{N}\sum_{k=1}^{K}[\log(w_k)P(Y_j \in M_k \mid X_j, \Theta^{(t)})] + \lambda[\sum_{k=1}^{K} w_k - 1] \tag{6}$$

A partial derivative on Eq. (6) by w_k, we have that,

$$\frac{\partial J_w}{\partial w_k} = \sum_{j=1}^{N}[\frac{1}{w_k} P(Y_j \in M_k \mid X_j, \Theta^{(t)})] + \lambda = 0 \tag{7}$$

$$w_k^{(t+1)} = \frac{\sum_{j=1}^{N} P(Y_j \in M_k \mid X_j, \Theta^{(t)})}{N} \tag{8}$$

Secondly, to optimize θ_k by,

$$J_\theta = \sum_{j=1}^{N}\sum_{k=1}^{K}\log(f_k(X_j \mid Y_j \in M_k, \theta_k))P(Y_j \in M_k \mid X_j, \Theta^{(t)}) \tag{9}$$

For Gaussian distribution

$$f_k(X_j \mid Y_j \in M_k, \theta_k) = \frac{1}{\sqrt{(2\pi)^d \mid \sum_k \mid}} e^{[-\frac{1}{2}(X_j-\mu_k)^T \sum_k^{-1}(X_j-\mu_k)]} \tag{10}$$

Where T is transpose matrix, -1 is an inverse operator on matrix. Σ is covariance on matrix X_j (signal). By the same Lagrange operator we get,

$$\sum_k^{(t+1)} = \frac{\sum_{j=1}^{N} P(Y_j \in M_k \mid X_j, \Theta^{(t)})(X_j - \mu_k)^2}{\sum_{j=1}^{N} P(Y_j \in M_k \mid X_j, \Theta^{(t)})} \tag{11}$$

And

$$\mu_k^{(t+1)} = \frac{\sum_{j=1}^{N} X_j P(Y_j \in M_k \mid X_j, \Theta^{(t)})}{\sum_{j=1}^{N} P(Y_j \in M_k \mid X_j, \Theta^{(t)})} \tag{12}$$

3.2 Feature extraction

3.2.1 EEG features

(1) Self-Similarity (SSIM)

The common time-domain characteristics of electrical signals can be used in the calculation of EEG signals and have better recognition efficiency, such as the average of the absolute of the first-order difference, for N segment signals $X(t)$, the self-similarity feature is defined by,

$$ssim = \frac{\frac{1}{N-1}\sum_{n=1}^{N-1} \mid X(n+1) - X(n) \mid}{\sigma_X} \tag{13}$$

(2) Energy (E)

The activity of the cerebral cortex affects the amplitude of EEG, which in turn reflects energy fluctuations,

$$E_X = \sum_{n=1}^{N} \mid X(n) \mid^2 \tag{14}$$

(3) Complexity (C)

$$C = \sqrt{\frac{\frac{1}{N}\sum_{n=1}^{N}(X(n)-\mu_X)^2}{\frac{1}{N}\sum_{n=1}^{N}(X'(n)-\mu_X)^2}} \tag{15}$$

Where $X'(n)$ is derivative of $X(n)$, μ_X is an expectation of matrix X (signal).

(4) High order crossing (HOC)

Covert EEG signal to $Z(n)$ with 0 mean. K is the number of filters, construct a series firstly as,

$$L_k\{Z(n)\} = \sum_{j=1}^{k}\frac{(k-1)!}{(j-1)!(k-j)!}(-1)^{j-1}Z(n-j+1) \tag{16}$$

And the threshold function is assigned by,

$$X_n(k) = \begin{cases} 1, L_k\{Z(n)\} \geq 0 \\ 0, L_k\{Z(n)\} < 0 \end{cases} \tag{17}$$

Where, $k = 1, 2, \ldots, K$; $n = 1, 2, \ldots, N$, then we have that,

$$HOC =_k \sum_{n=2}^{N}(X_n(k) - X_{n-1}(k))^2 \tag{18}$$

(5) Discrete Fourier Transfer (DFT)

$$DFT(X(n)) = \sum_{n=1}^{N-1}X(n)e^{-j(\frac{2\pi}{N})nk} \tag{19}$$

Where, $W_N = e^{-j(\frac{2\pi}{N})}$ is transform matrix.

(6) Power Spectral Density (PSD)

PSD is the distribution of signal energy with frequency and defined as,

$$PSD = \int_0^{+\infty} G(\omega)d\omega \tag{20}$$

Where power $G(\omega)$ is the average power of frequency ω.

(7) Short-Time Fourier Transform (STFT)

STFT can improve the robustness to noise, use the window function of equal length, calculate the Fourier transform in the window, and here Gaussian function is used for windowing, in window $\omega(n-t)$, we have that,

$$F(n,\omega) = \int_{-\infty}^{+\infty}\omega(n-t)X(t)e^{-j\omega t}dt \tag{21}$$

For current emotion recognition tasks, the t window is 1-2 seconds.

3.2.2 Eye-tracking signal features

Many studies on eye movement signals have proved that the corneal part of the eyeball is a positive electrode, and the retina part is a negative electrode. This retinal electrostatic potential signal existing between the retinal pigment epithelium and the photoreceptor cells is called an electrooculogram, namely the EOG signal. When the eyes are not moving forward, a stable reference potential can be recorded. For every 1° movement of the eyes in the horizontal or vertical direction, voltages of about 16uV and 14uV will be generated, respectively; because the eye electrical signal is relative to EEG in terms of the signal, the amplitude is relatively strong, about 15-200uV, so the signal can be directly detected by the electrodes placed around the eye socket. In the E.T. signal, this research also used time-domain and frequency-domain based features as follows:

(1) Electrooculography power density estimation ((EOG-PDE). For eye-tracking signals, the frequency domain method mainly uses EOG power spectrum estimation. It was divided into non-parametric estimation and parameter estimation. Non-parametric estimation is established on the Fourier transform, including the autocorrelation method, periodogram method, and so on. The parameter estimation includes the power spectrum estimation method of the A.R. model. In this study, a parametric estimation method was used. If the autocorrelation function of an eye movement signal $X(n)$ is $r(k)$, then according to Eq. (20), its power spectral density is written as,

$$P(w)=\sum_{k=-\infty}^{+\infty} r(k)e^{-jwk} \tag{22}$$

Where $r(k)=E[X(n)X^{*}(n+k)]$ * is a conjugation operator on matrix. $E(\cdot)$ is an expectation.
By periodogram method, and by Eq. (22), we have the estimation on $P(w)$,

$$\hat{P}(w)=\frac{1}{N-k}\sum_{k=-(N-1)}^{N-1} X(n)X^{*}(n-k)e^{-jwk} \tag{23}$$

Where, $0\le k\le N-1$.
The autocorrelation estimation is adopted because the estimated value is proportional to the power of the sinusoidal signal, which has high computational efficiency. For the m-th windowed, zero-padded frame of the eye-tracking signal $X(t)$, we have that,

$$X_m(n)\ @w(n)X(n+mR) \tag{24}$$

Where, $n=0,1,...,M-1$, $m=0,1,...,K-1$; R is window hop size; K is the number of frames. Then the periodogram of the m is given by,

$$P_{X(m),M}(\omega_k)=\frac{1}{M}|FFT_{N,K}(X_m(n))|^2@\frac{1}{M}|\sum_{n=0}^{N-1}X_m(n)e^{-j\frac{2\pi nk}{N}}|^2 \tag{25}$$

(2) Center gravity frequency (CGF), for eye-tracking signal $X(t)$, the frequency of the signal is f, we have that,

$$CGF=\frac{|\sum_{i=1}^{N}P(f_i)\times f_i|}{|\sum_{i=1}^{N}P(f_i)|} \tag{26}$$

Where $P()$ is from Eq. (22)
(3) Frequency variance (F.V.)

$$FV=\frac{\sum_{i=1}^{N}P(f_i)\times f_i^{2}\sum_{i=1}^{N}P(f_i)-|\sum_{i=1}^{N}P(f_i)\times f_i|^2}{|\sum_{i=1}^{N}P(f_i)|^2} \tag{27}$$

(4) Root mean square frequency (RMSF)

$$RMSE=\frac{1}{N}\sqrt{\sum_{i=1}^{N}X_i(t)-\mu_i|^2} \tag{28}$$

Where, μ_i is the i-th mean.
(5) Power spectral entropy (PSE). PSE is measuring of signal irregularity, which sums the normalized signal spectral power.

$$PSE=-\sum_{i=1}^{N}\frac{P(w_i)}{\sum_{k=1}^{K}P(w_k)}\ln\left[\frac{P(w_i)}{\sum_{k=1}^{K}P(w_k)}\right] \tag{29}$$

So, we defined seven features of EEG and five features for eye-tracking signals. By combing those 12 features, the normalization operator may be defined as in the next section.

3.2.3 Normalization

Normalization is a data preprocessing technique that is widely used in pattern recognition. EEG signal or E.T. signal is a very complex non-stationary and nonlinear signal, and the frequency components contained in different moments are different. The normalization method takes the maximum, minimum, median, mean, and other statistically

significant values. When the eigenvalues differ significantly, these statistical values are also greatly affected by the larger eigenvalues. The normalized values, the difference between the characteristic values, is still significant. When the entire data set or the whole subject is a single normalized data range, the considerable difference in the feature value between the issues and the significant difference in the feature value between the electrodes will influence emotional factors challenging to highlight. The result is not ideal. Suppose the characteristic value corresponding to a single characteristic attribute of a single subject is used as a single normalized data. In that case, there will be no mutual influence of typical values between subjects and electrodes. The normalized results obtained will be better to keep the internal characteristics of the original data. Here the max-min method is used for processing fused EEG and E.T. signals.

$$\mathrm{l}(k)=\frac{X(k)-\min(X(n))}{\max(X(n))-\min(X(n))} \tag{30}$$

Where, $k=1,2,...,n$;

3.3 Deep neural networks

3.3.1 Fundamental structures

A deep neural network is a series of new structures and methods evolved to make multilayer neural networks with more layers trainable and able to work. The concepts of partial receptive field and sharing of powers are used to reduce network parameters significantly. The key is that this structure is very suitable for visual tasks and similar to the human brain's working principle. In a broad sense, the network structure of a deep neural network is also a multilayer neural network. The traditional multilayer neural network has only the input layer, hidden layer, output layer. The number of hidden layers is determined according to needs, and there is no exact theoretical derivation to explain how many layers are appropriate (shown in Figure 2).

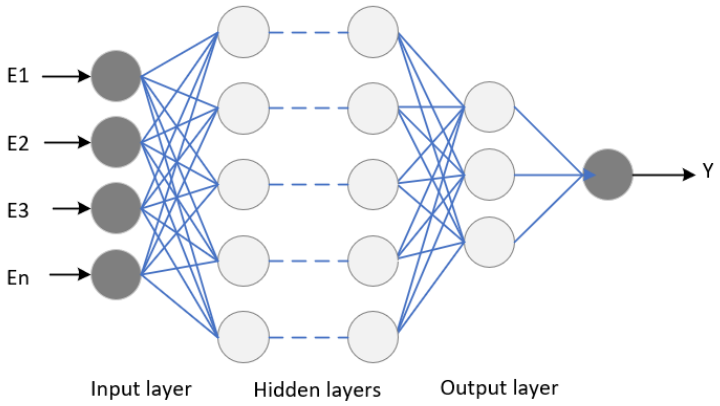


Figure 2. A typical deep neural network

The deep neural network is a discriminative model, which can be trained using a backpropagation algorithm [37]. The weight can be updated by the stochastic gradient descent method (SGD) as,

$$\Delta w_{ij}(t+1)=\Delta w_{ij}(t)+\eta\frac{\partial C}{\partial w_{ij}} \tag{31}$$

Where, η is the learning rate and C is the cost function, the choice of this function is related to the type of learning and the activation function.

3.3.2 Designed deep gradient neural network

The propagation algorithm: for each sample, the object is $J(W,b,x,y)=\frac{1}{2}\|a^L-y\|_2^2$, and the L-th layer of fused deep gradient neural networks (DGNNs), weights, and bias of W and b satisfied the following equation.

$$a^L=\sigma(W^L a^{L-1}+b^L) \tag{32}$$

The lost function for the output layer is,

$$J(W,b,x,y)=\frac{1}{2}\|\sigma(W^L a^{L-1}+b^L)-y\|_2^2 \tag{33}$$

Deviation by weights, we have that,

$$\frac{\partial J(W,b,x,y)}{\partial W^L}=(a^L-y)(a^{L-1})^T \text{ e } \sigma'(z) \tag{34}$$

$$\frac{\partial J(w,b,x,y)}{\partial b^L}=(a^L-y) \text{ e } \sigma'(z^L) \tag{35}$$

Where e is Hadamard product? Then the W^L b^L and can be inducted hereafter by using Algorithm 1.

Algorithm 1: the weights and bias of DGNN

Required: initial W, b, a_l: the step length of the l-th layer, L: total number of layers, epsilon: threshold for loop, M: max iterations.

Output: layers W and b

```
Initialization W and b
FOR iter in 1 TO M
    FOR i in 1 TO m
        INPUT DGNN by x_i←a_l
        FOR l in 1 TO L
            FORWARD: a_i,l←segma(W_la_i,l-1+b_l)
            CALCULATE loss function by delta_i,l
        ENDFOR
        FOR l in L TO 2
            BACKWARD: delta_i,l←(W_(i+1))delta_i,(l+1)
        ENDFOR
    ENDFOR
    FOR l in 2 TO L
        UPDATE the l-th layer's W_l and b_l
    ENDFOR
    IF the change of W, b < epsilon THEN
        BREAK
    ENDIF
ENDFOR
RETURN
    W, b
```

Based on the fundamental framework of deep neural network, an optimization on gradient was proposed in this research. During the neural network training process, the parameters of each layer will continue to change as the iteration progresses, which will cause the distribution of the input data of the subsequent layer to continue to change. This problem is called internal covariate shift. Each layer needs to adapt to the input data distribution during training, which requires us to adjust the learning rate in the iterative process and acceptable initialization weight parameters. To solve this problem, we need to normalize each layer of the neural network's input data. One solution is Batch Normalization, a particular layer in the network used to batch normalize the input data of the previous layer and then send it to the next layer for processing. It may speed up the training process of the neural network [38].

In addition to the standard mini-batch stochastic gradient descent method. They have achieved better results in many experiments and practical applications. AdaGrad is an adaptive gradient, that is, an adaptive gradient algorithm. The most immediate improvement of the gradient descent method. The only difference is that AdaGrad adjusts the learning rate based on the historical gradient values of the previous iterations. The AdaDelta algorithm is also a variant of the gradient descent method, and the gradient value is also used to construct the updated value of the parameter in each iteration. Adam algorithm (adaptive moment estimation) creates two vectors m and v from the gradient term, and their initial value is 0. Like the standard gradient descent method's weight update formula, the NAG algorithm constructs a vector v with an initial value of 0. Parameter initialization and momentum terms are both critical to the convergence of the algorithm. The initial weight and momentum terms are crucial for the training optimization problem of deep neural networks and cyclic neural networks, and both are indispensable [39]. If the initial value is set improperly, it

will be difficult to converge to a good effect even if the momentum term is used; on the other hand, if the initial value is set well, but the momentum term is not used, the convergence effect will be compromised.

SGD is the most commonly used for optimization. According to the data generation distribution, randomly sample mm small batches (independent and identically distributed) samples and obtain an unbiased estimate of the gradient by calculating the mean of their gradients shown in Algorithm 2.

Algorithm 2. Improving gradients of DGNN using randomly descent

REQUIRED: $\varepsilon(k)$: learning rate; θ : initial parameters

WHILE stop rule **DO**

Extract minimum batch from signal samples of [x(1), x(2), …, x(m)]

Estimate the gradient by

$$g \leftarrow g + \frac{1}{m}\nabla\sum_i L(f(x(i);\theta), y(i))$$

Update: $\theta \leftarrow \theta - \varepsilon g$

END WHILE

SGD is the most popular optimization algorithm, but its learning process is sometimes a bit slow. The momentum method is designed to speed up the learning process. From a formal point of view, the momentum algorithm introduces the variable v to act as a speed role-representing the direction and rate of movement of the parameter space parameter. The speed is considered the exponential decay average of the negative gradient. The introduction of the term momentum is also to compare the concept of power in physics. The hyperparameter αα determines how fast the previous gradient contribution decays. The update rule is as follows:

$$v = \alpha v - \varepsilon\nabla_\theta(\frac{1}{m}\sum_{i=1}^{m} L(f(x(t);\theta), y(t))) \tag{36}$$

SGD and its related mini-batch, or a more generalized gradient-based online learning algorithm, an important property is that the calculation time of each step of the update does not depend on the total number of training samples. When the full data set is large, it can also converge, and SGD often converges within an acceptable error range before processing the entire training set. Research on the algorithm's convergence rate generally measures the excess error $J(\theta) - \min(J(\theta))$, that is, the amount by which the current cost function exceeds the lowest possible cost. When SGD is applied to convex problems, the magnitude of the error after k iterations is $O(k)$, and the magnitude of the error is $O(k)$ in the case of strong convexity. Without additional assumptions and auxiliary information, the above boundaries cannot be further improved. In theory, batch gradient descent SGD has a better convergence rate. However, some scholars have pointed out that the generalization error will not decrease faster than $O(k)$, so it is not worth exploring convergence for machine learning algorithms. The optimization algorithm is faster than $O(k)$, because often too fast convergence corresponds to overfitting. SGD only needs a small number of samples to calculate the gradient for large data sets to achieve rapid initial updates. However, due to the advanced analysis that SGD loses a constant multiple of $O(k)$, it can gradually increase the batch size of small batches during learning to weigh and make full use of both batch stochastic gradient descent advantages. Algorithm 3 is improved by velocity based on Algorithm 2.

Algorithm 3. Improving gradients by velocity

REQUIRED: ε : learning rate; θ : initial parameters; v : initial velocity

WHILE stop rule **D.O.**

Extract minimum batch from signal samples of [x(1), x(2), …, x(m)]

Estimate the gradient by

$$g \leftarrow g + \frac{1}{m}\nabla\sum_i L(f(x(i);\theta), y(i))$$

$$v \leftarrow \alpha v - \varepsilon g$$

Update: $\theta \leftarrow \theta + v$

END WHILE

Continuously, we have:

$$v = \alpha v - \varepsilon \nabla_{\theta} (\frac{1}{m} \sum_{i=1}^{m} L(f (x(t); \theta + \alpha v), y(t))) \tag{37}$$

The parameters α ε and play a similar role to the standard momentum method. The difference between Nesterov momentum and standard momentum lies in the calculation of the gradient. In Nesterov momentum, gradient calculation after applying the current speed can be understood as Nesterov momentum adds a correction factor to the standard momentum method [40]. In the case of the convex optimization problem using batch gradient descent, Nesterov momentum increases the additional error convergence rate after kk steps from $O(k)$ to $O(k^2)$, which does not improve the convergence rate for SGD. So, if we continuously combine the complete Nesterov momentum and the improved gradient introduced in Algorithm 1, 2, and 3, then the grad computing for optimization of the fused deep neural network is finally formalized in Algorithm 4.

Algorithm 4. Improving gradients by velocity

REQUIRED: ε : learning rate; θ : initial parameters; v : initial velocity; δ : constant

WHILE stop rule **D.O.**

Extract minimum batch from signal samples of [x(1), x(2), …, x(m)]

Initial

Estimate the gradient by

$$\hat{\theta} \leftarrow \theta + av$$

$$g \leftarrow g + \frac{1}{m} \nabla \sum_{i} L(f (x(i); \theta), y(i))$$

Square gradients accumulate by

$$r \leftarrow r + g \text{ e } g$$

$$\Delta\theta \leftarrow \frac{\varepsilon}{\delta + r} \text{ e } g$$

$$\theta \leftarrow \theta + \Delta\theta$$

$$v \leftarrow \alpha v - \varepsilon g$$

$$\theta \leftarrow \theta + v$$

END WHILE

Where e is a product between vectors? Optimization is the process of continually searching for parameters to minimize or maximize the objective function. The core problem of deep learning is a complicated optimization problem. Therefore, in the decades after introducing neural networks, the difficulty of deep neural networks' optimization problems is an essential factor preventing them from becoming mainstream. By running the classic gradient descent optimization method, you can get a good enough local minimum, allowing us to make significant progress on many common problems. When building a neural network model, we choose the best optimizer to quickly converge and learn correctly while adjusting internal parameters to minimize the loss function. This type of data processing method of deep learning can be seen as the classic signal processing system's evolution. By learning and optimizing this type of data-driven approach, the algorithm relies on fewer assumptions about the data and becomes more flexible and scalable. The contribution of this paper is to analyze, preprocess and interactively annotate signals, extract features and convert signals, train deep neural networks, and build deep learning models for real application scenarios such as biomedicine, audio, communications, and radar.

4. Results and discuss

4.1 Experimental design and signal processing

4.1.1 Emotion event stimuli

Visual stimulus is the most commonly used emotion induction method, that is, to present the subjects with emotionally coloured text, pictures, and other stimulating materials to induce the subjects' target emotions. Currently, visual stimulation has A relatively complete library of standard stimulus materials has been formed. In terms of text, the English emotional word system launched by the National Institute of Mental Health (NIMH), (Affective Norms for English Words (ANEW) [41] and English Affective Norms for English Text (ANET) [42] are widely recognized text dynamic stimulation material libraries [43]; in terms of pictures, NIMH established the International Affective Picture

System (International Affective Picture System) provides more options for emotion-induced research [44]. In this emotion event stimuli experiment, the stimuli system is IAPS; basic emotions are anger, disgust, fear, sad, expect, happiness, surprised, and trust [45].

4.1.2 Eye-tracking experiment

The experiment selected five males and five females with the normal eye function, aged between 23-45. Participants received the investigation alone in a quiet and bright laboratory. Participants sat 60 cm away from the eye tracker and keyboard and found a comfortable posture to minimize head movement. Once the message is consistent, the experimenter will prompt the subjects' eyes to calibrate the eye tracker and ask their eyes to follow the white screen's blue circle. Participants were told that they would see 32 pictures, and their task was to watch these 32 pictures freely. Two facial images with different emotions will appear on the screen at the same time for 5 seconds. In the subject's field of vision, the only thing that changed was the picture's content. The eye movements are recorded using a remote video eye-tracking system, which uses a single eye pupil to track at 500 Hz. To calibrate the eye position, a 13-point grid is used, and the calibration process is repeated until the average error is less than 0.5°. A speed threshold of 30°/s, an acceleration threshold of 8000°/s2, and a minimum deflection threshold of 0.1° are used to define the saccade. In this interval, the amplitude of the recorded eye movement exceeding six relative speed thresholds based on the median, standard deviation for a duration of at least six samples (12 ms) is between 0.1° and 1°. Micro saccade detection is only performed in gaze tests that do not include saccades greater than 1° according to the criteria described in Eye-Tracking. The parameters of the E.T. signal acquisition are following by 600Hz sampling frequency, 0.3° at optimal conditions of accuracy, 0.06° RMS at optimal conditions of precision, 0.01° RMS at optimal conditions of filtered precision, 30° maximal gaze angle, 34 cm x 26 cm dimension, 55 to 75 cm from the eye tracker reference point of operating distance, less than three frames of total system latency, less than 150 ms gaze recovery time, and 100 μs tracker and client time synchronization (Shown in Figure 3).

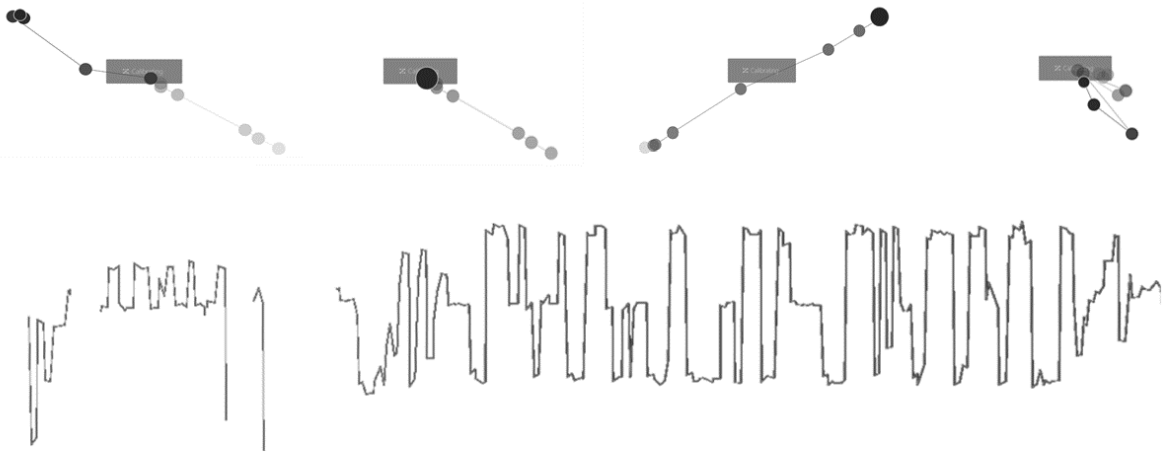


Figure 3. eye movement and eye-tracking signals

4.1.3 EEG signal acquisition

The 64-channel active electrode system with a D.C. amp amplifier was used to record EEG data (Shown in Figure 4). Place 61 electrodes on the scalp equidistantly, one on the forehead (approximately 25 mm above the nostril), and the other two on the left and right infraorbital rim. The impedance of all electrodes is reduced to below 5kΩ. The data were recorded at a sampling rate of 1000 Hz, and the online bandpass was filtered between 0.016 Hz and 250 Hz. During the recording, all electrodes are referenced to the tip of the nose.

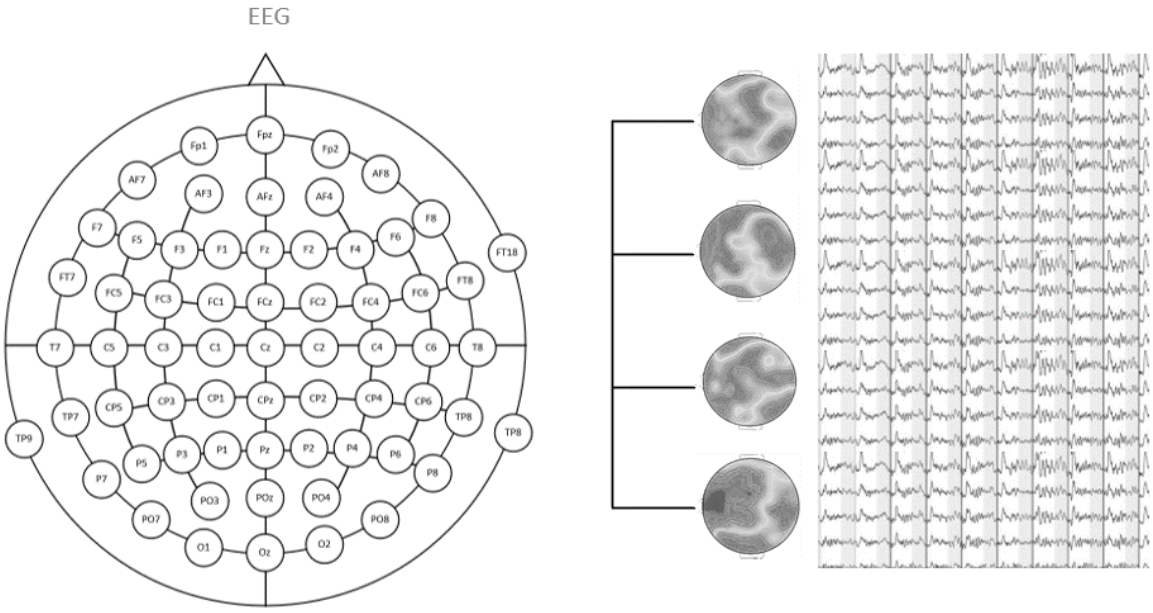


Figure 4 64-channels-61 electrodes of EEG on the distribution of cerebral cortex electrodes and their data acquisition, including imaging and waveform diagrams

4.1.4 Stimuli and signal synchronization processing

In the experiment, 8 groups of emotional stress pictures were used, each of which was two pictures of the same gender and different emotions, set emotion labels to [E1, E2, …, E8] = [angry, disgust, fear, sad, expect, happy, surprised, trust], in which E1 to E4 are negative, and E5 to E8 are positive. The negative, neutral, and positive emotion pictures, including all emotion classification, were randomly presented one by one in the eye tracker. The researchers designed four groups of images containing positive and negative emotions, and the other four groups were random combinations of positive and negative pictures. After that, the fixation time of the subjects in these eight groups was calculated. Participants with positive emotions avoided watching negative emotions pictures and looked at positive emotions for much longer. Similarly, issues with negative emotions are more likely to pay attention to negative images. The software directly sends the TTL signal to the Trigger-in interface of the multi-parameter synchronizer of the EEG device (the multi-parameter synchronizer is wirelessly synchronized with the EEG device), so that the eye movement signal may trigger the mark in the EEG; the accuracy of the software synchronization depends on the corresponding eye and the sampling frequency of the actuator. For example, a glasses eye tracker with a sampling rate of 100Hz can achieve a synchronization accuracy of 10ms; a desktop eye tracker with a sampling rate of 600Hz can achieve a synchronization accuracy of <2ms. The spectrum is directly connected to the EEG device through the DB9 port. A level signal is sent to the DB9 port through the Trigger-out interface of the multi-parameter synchronizer to achieve triggering. According to the Spectrum sampling rate (300/600/1200Hz), the highest synchronization accuracy can be achieved. 1ms. The sampling frequency is 50/100Hz, so the synchronization accuracy of 10/20ms is achieved ((Shown in Figure 5).

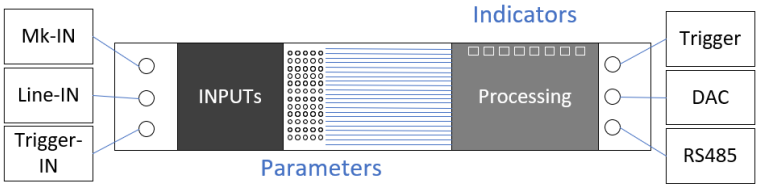


Figure 5 signal synch with multiple parameters

4.2 Classification and performance

4.2.1 Results

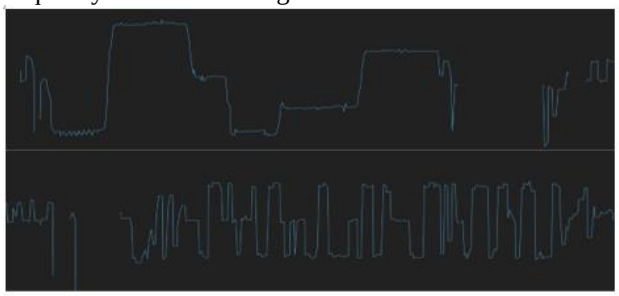
The first step is to convert the eye-tracking signal to plain text using IDF (a raw format of eye-tracking signal) convert and then construct a MATLAB data format for further preprocessing. EEG-Lab and EYE_EEG tools were used for

preprocessing the signals [46, 47]. An eye-tracking matrix was acquired through a tracking system and processed by MATLAB2020a, as shown in Table 1.

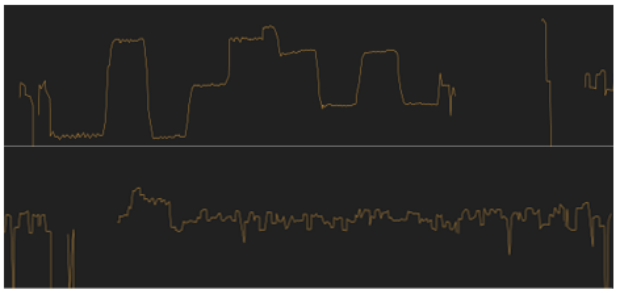
Table 1. eye-tracking matrix with event stimuli under eight emotions picture IAPS system (partial)

Frequency	Sampling	GP_x	GP_y	Speed	Peak	TSA	Amplify	RMS	IF	SNR	THD	Affective
4.7	3.9	1.2	0.1	3.0	3.4	4.1	5.1	3.2	1.3	0.1	3.3	'expect'
3.9	3.2	1.3	0.3	3.0	3.3	4.5	4.5	3.1	1.2	0.1	3.3	'expect'
5.1	3.1	1.2	0.3	3.2	3.2	5.1	5.1	3.1	1.2	0.2	3.4	'expect'
4.2	3.1	1.2	0.3	3.3	3.2	5.3	4.9	3.2	1.2	0.1	3.2	'expect'
4.7	3.1	1.3	0.2	3.3	3.2	5.2	4.5	3.1	1.2	0.2	3.4	'expect'
5.1	3.2	1.2	0.3	3.4	3.2	5.1	4.9	3.2	1.3	0.2	3.2	'expect'
4.2	3.3	1.4	0.2	3.2	3.6	4.9	5.1	3.3	1.4	0.3	3.5	'expect'
4.5	3.1	1.6	0.1	3.3	3.2	4.9	5.2	3.1	1.2	0.2	3.5	'expect'
3.9	3.6	1.5	0.1	3.6	3.3	5.2	5.3	3.2	1.4	0.1	3.2	'expect'
5.6	2.9	1.2	0.3	3.7	3.4	5.2	5.3	3.7	1.4	0.2	3.5	'expect'
5.7	3.9	1.3	0.1	3.6	3.2	5.1	5.3	3.3	1.2	0.2	3.5	'expect'
3.2	3.2	1.7	0.3	3.3	3.4	5.2	5.2	3.4	1.3	0.2	3.3	'expect'

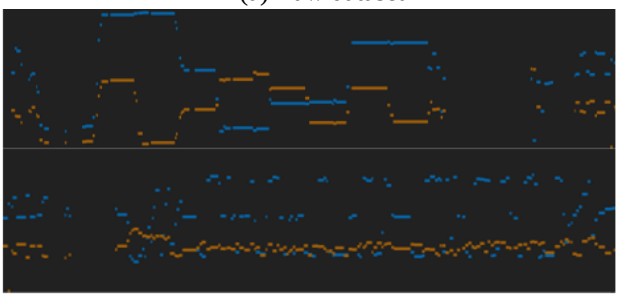
In the table, GP_x and GP_y are gaze position coordinators; Gaze is a process to estimate and track the 3D line of sight of a person, or simply, where a person is looking. TSA is a signal with Tyramine signal amplification. RMS means root mean square. IF stands for intermediate frequency; and SNR is signal-to-noise ratio. THD total harmonic distortion. The raw waveform-based signal, Gaze coordinators of X and Y, and feature calculation of PSD and time-frequency are shown in Figure 6.



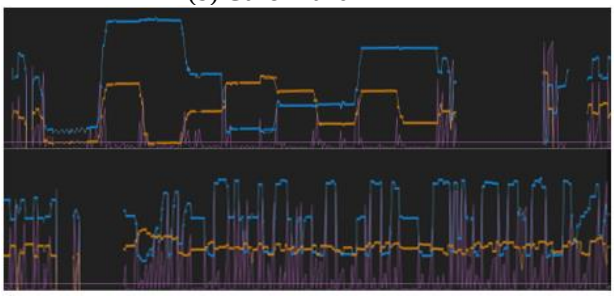
(a) Raw dataset



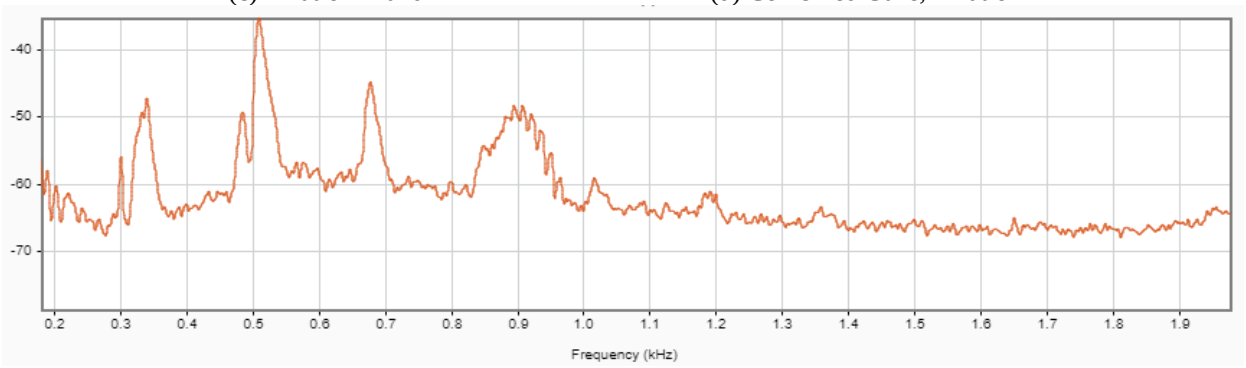
(b) Gaze X and Y



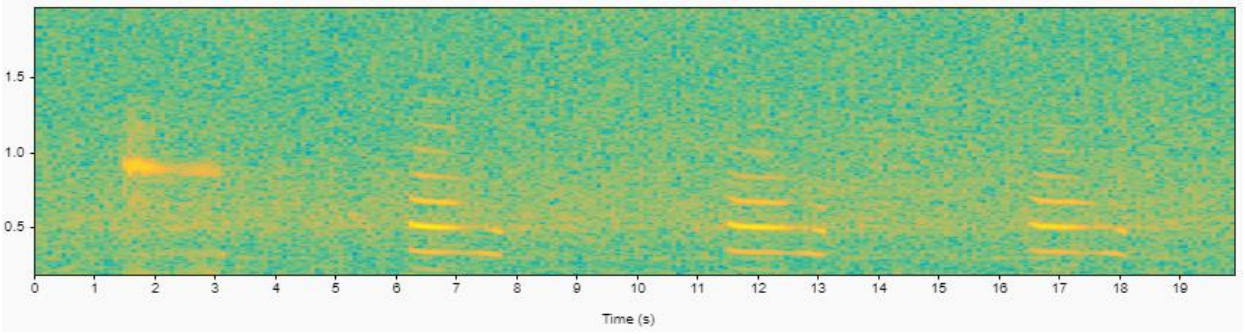
(c) Fixation X and Y



(d) Combined Gaze, Fixation



(e) feature of PSD as an example for the eye-tracking signals



(f) Time frequency-based feature calculating for the eye-tracking signals

Figure 6. the raw dataset of eye-tracking signals and their preprocessing

And for EEG signals, Figure 7 is the plot of the frequency of EEG in partial event stimuli; and Figure 8 is the eight emotion classification stimuli by taking one as an example.

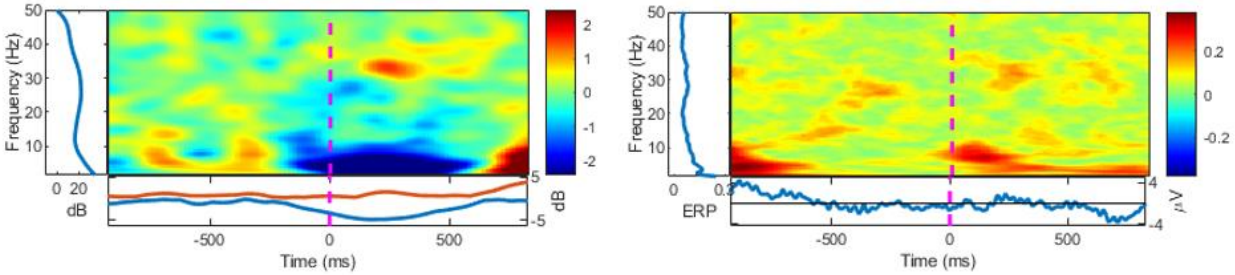


Figure 7. EEG frequency map plot

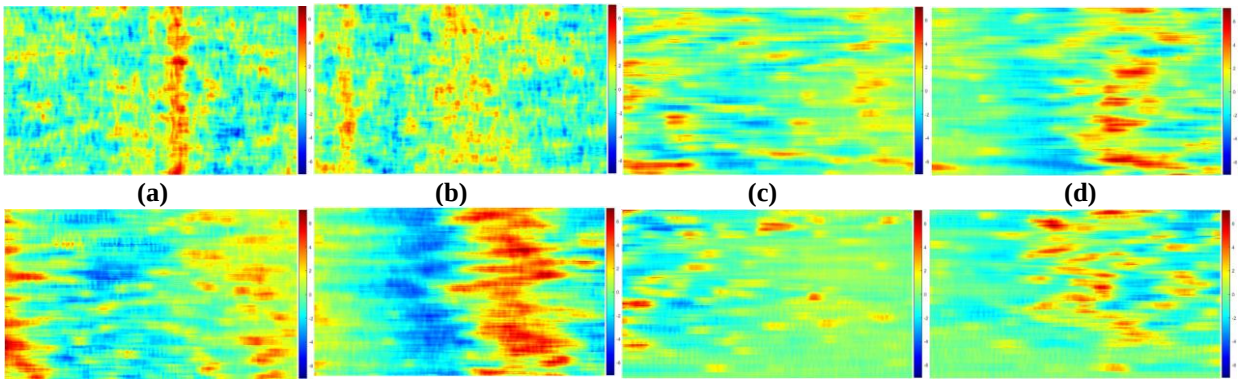


Figure 8. 8 stimuli EEG signals map

By taking one channel signal, Figure 9 shows the activate power spectrum (setting the global offset to -0.104), latency with potential volt distribution in the status of -607, component activities, and spectrum map with PSD. The waveforms of one polar mapping on brain EEG with feature map signals and all syn8-254 statuses, including 61 electrodes, were shown in Figure 10. After the signal acquisition and preprocessing, feature calculation is the next step; by using MATLAB2020a, all features were extracted, and for testing the stability of all features, a consistent matrix for the features is shown in Figure 11. By taking 15% for validation, 15% for testing, the proposed DGNN method outputs the predicted classes. The predictive matrix for the proposed DGNN method for classifying emotion event stimuli fused signals is shown in Figure 12.

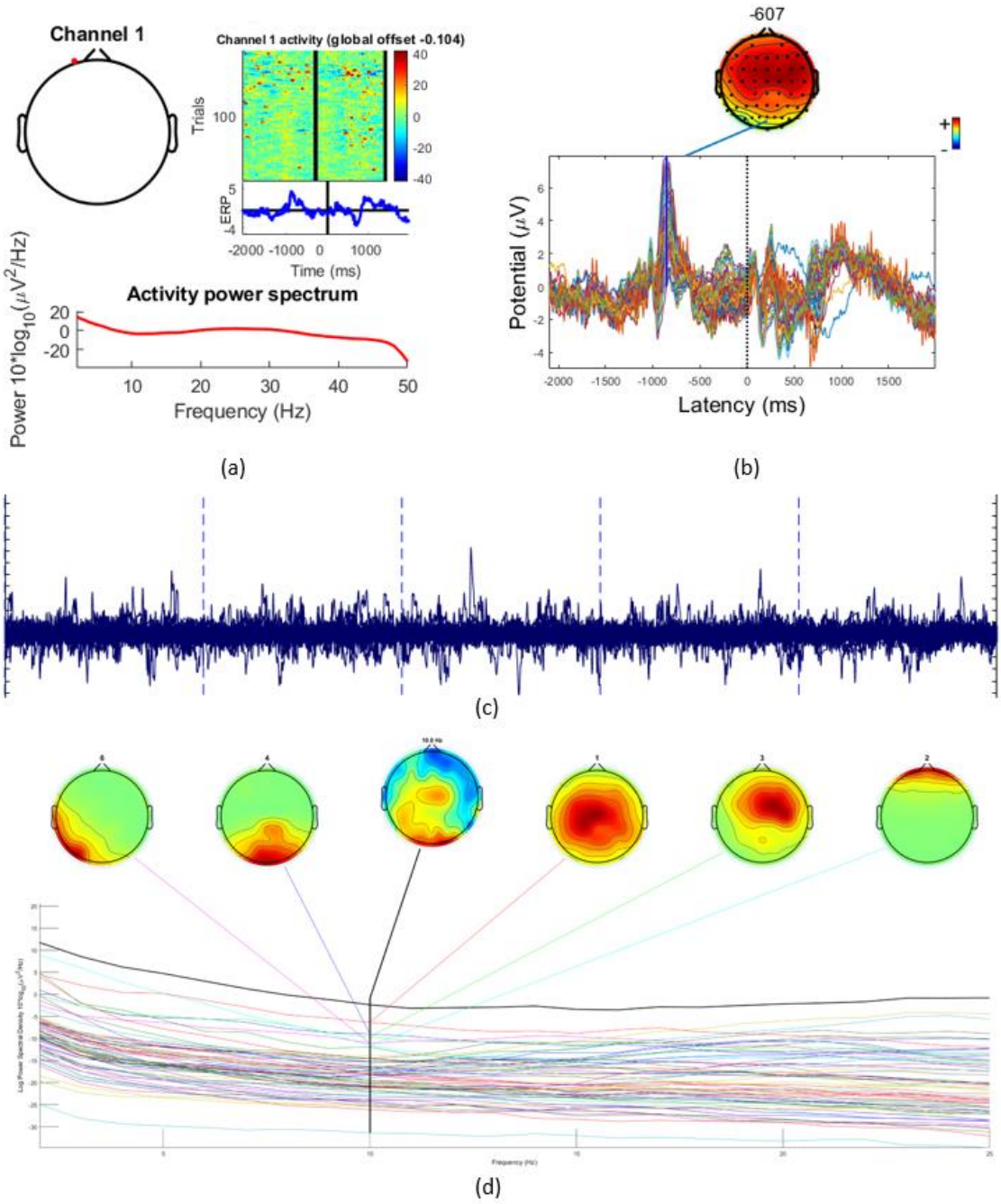


Figure 9. EEG processing with (a) Taking one channel for detecting activities with global offset -0.104 and their power spectrum, (b) potential with latency for the EEG signals, (c) component activate, and (d) spectrum map with PSD

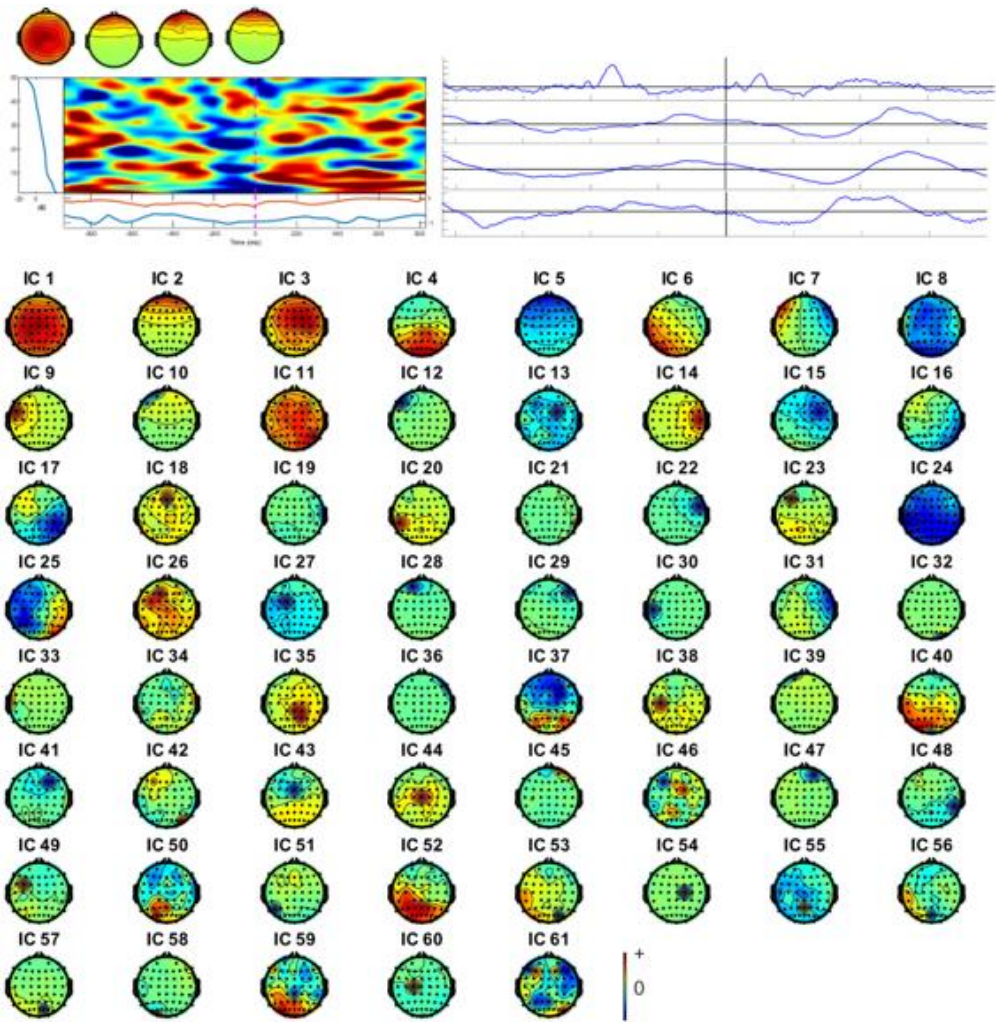


Figure 10. Waveforms of one polar mapping on the brain with feature map signals and all syn8-254 status

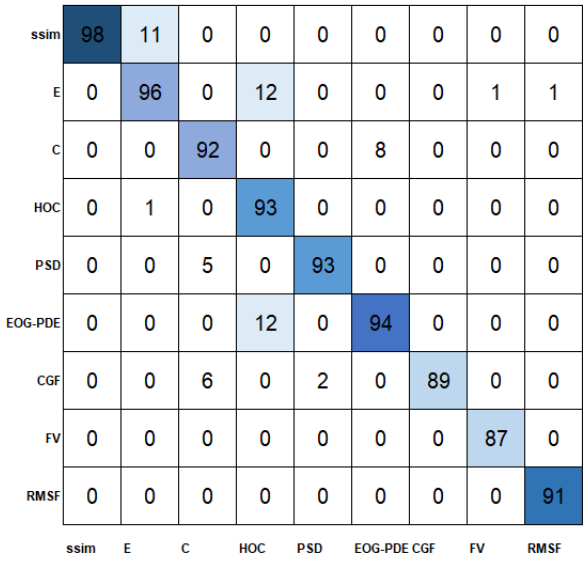


Figure 11. The consistency of the feature extraction for the fused signals.

Figure 12. DGNN method for classifying emotion event stimuli fused signals.

4.2.2 Discussion

Several neural networks were compared in this work, and firstly, a typical artificial neural network was designed for the fused signals. The structure was designed and illustrated in Figure 13. 15% of the inputs (organized emotion vectors including E.T. and EEG features and normalized by the proposed method in Section 2.2.3). Figure 14 shows the performance of the multiple layers artificial neural network for this case.

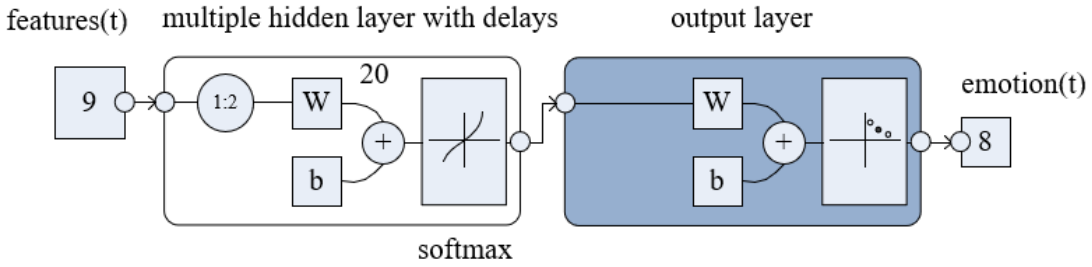


Figure 13. Multilayer artificial neural network for classifying fused E.T. and EEG signals

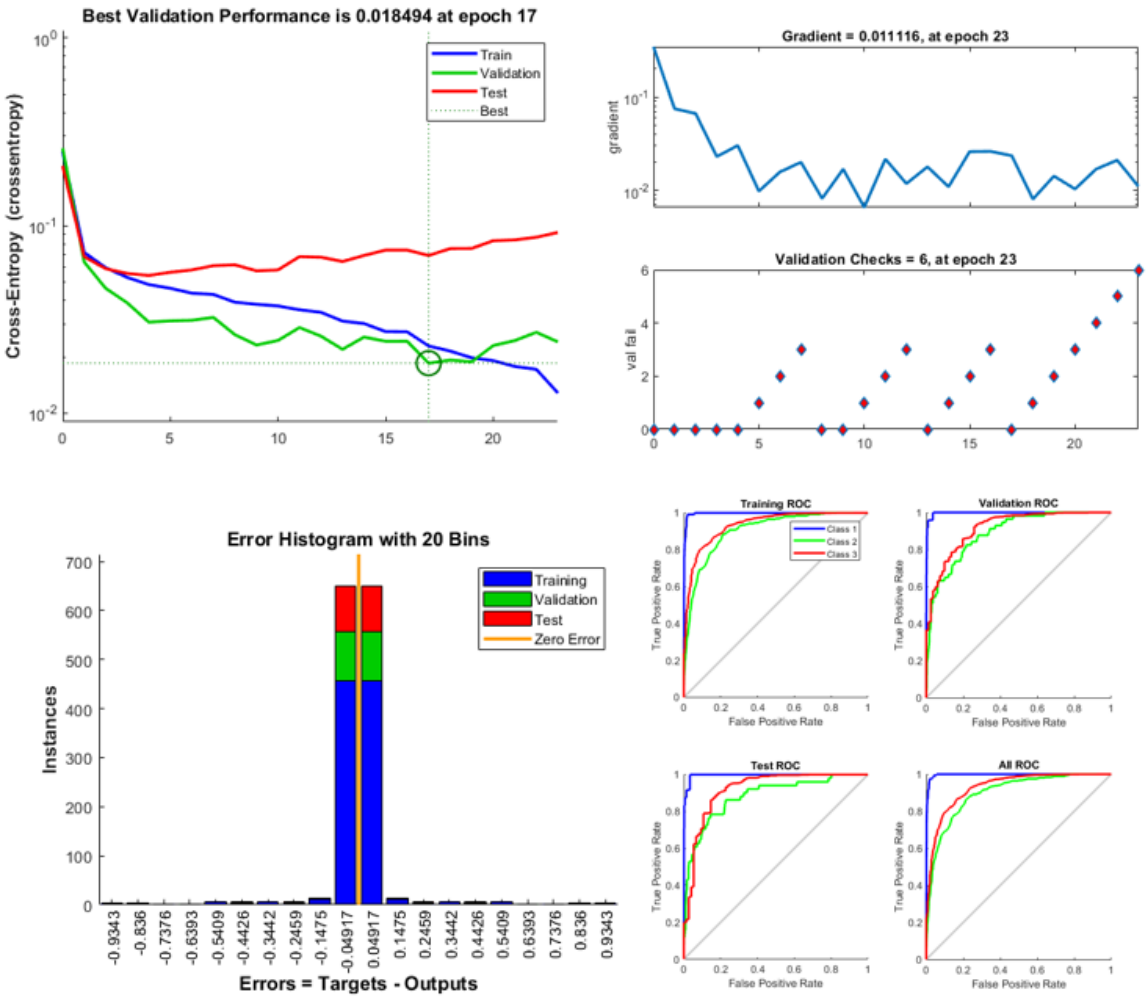
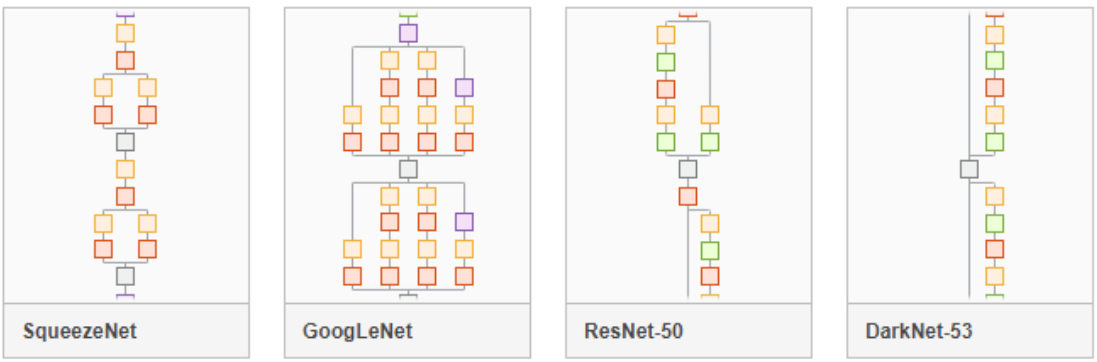


Figure 14. the regular ANN training and for fused prediction achieved lower accuracy.

Typical pre-trained deep neural networks for comparing analysis were selected, and the structure of the pre-trained neural networks are shown in Figure 15



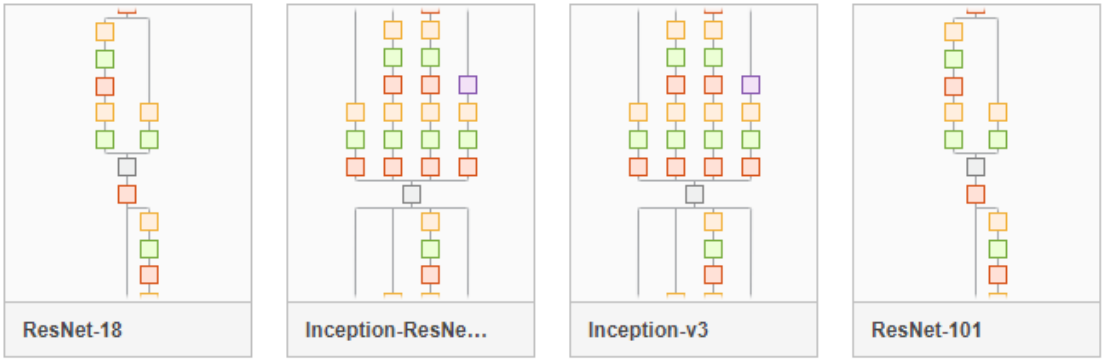


Figure 15. the structure of several deep neural networks

Now, some indices were selected for the evaluation of the comparing neural networks. The basic indices for signals are firstly introduced below:

(1) Basic indices. For the problem of single-label classification, the evaluation indicators mainly include Accuracy (Ac), Precision (Pr), Recall (Re), F-measurement (Fm), P.R. (Precision-Recall) curve, ROC (receiver operating characteristic curve), and AUC (Area Under Curve). Before calculating these indicators, we first calculate a few fundamental indicators. These indicators are based on two-category tasks and can also be extended to multi-category. The count label is a positive sample, and the number of positive samples classified as True Positive, referred to as T.P. The label is a positive sample, and the number of samples classified as unfavourable is False Negative, referred to as F.N. The label is a negative sample, and the number of positive samples classified as False Positive, referred to as F.P. The label is a negative sample, and the number of negative samples classified as True Negative, referred to as T.N. To determine whether it is a positive example, only a probability threshold T is required. If the predicted probability is greater than the threshold T, it is an upbeat class, and if the predicted probability is less than the threshold T, it is a negative class. The default is 0.5. If we reduce this threshold T, more samples will be identified as positive classes, which can increase the recall rate of positive classes, but at the same time, it will also bring more negative classes to be incorrectly classified as positive classes. If the threshold T is increased, the positive class's recall rate decreases, and the accuracy increases. If there are multiple categories, such as 1000 categories in the ImageNet1000 classification competition, the predicted type is the highest predicted probability. Accuracy is the probability that all samples are correctly classified, and different thresholds T can be used. $Ac = \frac{TP + TN}{TP + FP + TN + FN}$, $\mathrm{Pr} = \frac{TP}{TP + FP}$, $\mathrm{Re} = \frac{TP}{TP + FN}$, $Fm = \frac{2\,\mathrm{Pr}\,\mathrm{Re}}{P\,\mathrm{r} + R\,\mathrm{e}}$. Average precision is abbreviated as A.P., an indicator often used in retrieval tasks and regression tasks, which is equal to the area under the precision-recall curve.

(2) ROC curve and AUC indicator: the above accuracy rate accuracy, precision, recall rate Recall, F1 score, a confusion matrix is just a single numerical indicator. The classification algorithm's performance under different parameters may be performed by a curve, namely receiver operating characteristic (ROC). The ROC curve can be used to evaluate the performance of a classifier under different thresholds. In the ROC curve, each point's abscissa is the false positive rate (FPR), and the ordinate is the true positive rate (TPR), depicting the balance between the classifier's T.P. and F.P. The calculation of the two indicators is TPR=TP/(TP+FN), which represents the ratio of actual positive instances to all positive instances in the positive class predicted by the classifier; and FPR=FP/(FP+TN), which means the proportion of actual negative instances in the positive class predicted by the classifier to all negative instances. The larger the FPR, the more actual negative classes in the predicted positive class. Compared with the P.R. curve, the ROC curve has a good characteristic: when the distribution of positive and negative samples in the test set changes, the ROC curve can remain unchanged, that is, it is not sensitive to the imbalance of positive and negative samples. If the ROC curve of one classifier completely covers the other classifier, then it can be judged that the performance of the former is better. By quantitatively evaluating the performance of two classifiers through two ROC curves, the area under curve (AUC) indicator can be used; it is the area under the ROC curve, which represents a probability. AUC characterizes the likelihood that the predictive value of the positive sample is higher than the negative sample by the classifier. Of course, the premise is that the positive sample's predicted value should be higher than the negative sample.

(3) TAR (True Accept Rate): TAR represents the proportion of correct acceptance. Two images of the same person are taken multiple times, and the ratio of the similarity value exceeding the threshold T is counted. FRR (False Reject Rate) is the false rejection rate. The images of the same person are regarded as different people, and it is equal to 1-TAR. FAR (False Accept Rate) represents the percentage of false acceptance. Two images of different people are taken multiple times, and the portion of the similarity value exceeding "T" is counted. Increasing the similarity threshold "T", both FAR and TAR will decrease, which means that the proportion of correct acceptance and false acceptance will decrease, and the false rejection rate FRR will increase. Decreasing the similarity threshold "T", both FAR and TAR increase, the proportion of correct acceptance and the proportion of false acceptance both growth and the false rejection rate FRR decreases.

(4) Intersection-over-Union (IoU): IoU is the ratio of intersection and union. In target detection, it is defined as the ratio of the intersection of the area of two rectangular boxes and the union, $IoU = A \cup B / A \cap B$. If they overlap completely, IoU is equal to 1, which is the ideal situation. Generally, in a detection task, if IoU is greater than or equal to 0.5, it is considered to be recalled. If a higher IoU threshold is set, the recall rate decrease and the positioning frame will be more accurate. IoU is often used in image segmentation, and it does not have to be limited to the area of two rectangular boxes. For example, for two-class front background segmentation, then IoU= (real foreground pixel area $\cap$ predicted foreground pixel area)/ (real foreground pixel area $\cup$ predicted foreground pixel area), this index is usually more correct than directly calculating the classification of each pixel. The probability is low, and it is more sensitive to misclassification.

So, we finally evaluated the methods by using the indices defined above, and the results are listed in Table 2.

Table 2. Classifiers' performance under evaluation indices for the fused ET and EG signals.

Classifier Evaluation Metrics	**Ac**	**Pr**	**Re**	**Fm**	**PR**	**TPR_ROC**	**TAR_AUC**	**IoU**
ANN [48]	78.40%	81.23%	77.43%	78.12%	77.12%	7.11%	0.13%	0.91%
SqueezeNet [49]	81.40%	81.23%	77.43%	78.12%	77.12%	7.25%	0.13%	0.91%
G**oogleN**et [50]	81.76%	82.12%	76.43%	76.32%	75.32%	7.23%	0.12%	0.82%
ResNet-50 [51]	82.42%	82.33%	76.78%	79.34%	77.32%	7.23%	0.12%	0.87%
DarkNet-53 [52]	82.43%	84.65%	78.76%	81.32%	78.12%	6.31%	0.14%	0.88%
ResNet-18 [51]	85.54%	84.43%	82.45%	82.57%	81.11%	6.54%	0.12%	0.87%
Inception-ResNet [53]	85.43%	83.87%	84.67%	79.89%	82.81%	6.66%	0.14%	0.90%
Inception-v3 [53]	87.11%	84.65%	83.55%	84.90% (+)	81.78%	7.54%	0.14%	0.92%
ResNet-101 [51]	85.76%	82.48%	85.12% (+)	83.43%	82.32%	7.35%	0.14%	0.93%
DGNN (*)	88.10% (+)	85.98% (+)	82.69%	84.09%	84.87% (+)	7.55% (+)	0.14% (+)	0.94% (+)

* Proposed method in this research, (+) indicates better performance in the evaluation index.

5 Conclusion

Various physiological signals such as EEG, EOG, EMG, HVP, GSR, and skin temperature are used in emotional calculations, and a variety of equipment is required. In the experiment, it is necessary to minimize the noise in the signal acquisition process. The acquisition of EEG signals is more complicated, and EEG signals are very susceptible to interference from external factors. Therefore, experiments require a certain amount of time and effort. The ability to collect EEG signals and other physiological signals with high efficiency and high quality is the task of emotional computing. The preprocessing of the original physiological signal is also very important. Effective preprocessing can remove the original physiological signal's noise, improve the signal quality, and help feature extraction. In feature extraction, extracting appropriate features or merging different features will significantly impact the emotional computing model. On the other hand, due to the differences in physiology and psychology between different subjects, the emotions induced by different individuals may not be precisely the same for the same emotion-induced video, even if the same emotions are generated.

In future work, accurate extraction of signal features is an essential direction for improvement. There are many connections between EEG signals and other peripheral physiological signals and changes in human emotions. Deep learning can effectively learn the deep feature representations of samples and is very helpful in mining the emotional state information contained in physiological signals. However, different physiological signals have other feature extraction methods, such as EEG signals, and there are many types of feature extraction methods. How to extract appropriate features or merge different features will impact the emotional computing model.

References

[1] Geetanjali V K and Varsha H P, A study of vision based human motion recognition and analysis, International Journal of Ambient Computing and Intelligence, 7 (2016), 2, 75-92. https://doi.org/10.4018/IJACI.2016070104.

[2] Russell, J. A., A circumplex model of affect, Journal of Personality and Social Psychology. 39(1980),6, 1161-1178. https://doi.org/10.1017/S0954579405050340.

[3] Conte, H. R., & Plutchik, R. A circumplex model for interpersonal personality traits. Journal of Personality and Social Psychology. 40(1980), 4, 701-711. https://doi.org/10.1037/0022-3514.40.4.701.

[4]Magdin, M., & Prikler, F. Are instructed emotional states suitable for classification? demonstration of how they can significantly influence the classification result in an automated recognition system. IJIMAI. 5 (2019), 4, 141-147. https://doi.org/10.9781/ijimai.2018.03.002.

[5] Magdin, M., Sulka, T., Tomanová, J., & Vozár, M., Voice Analysis Using PRAAT Software and Classification of User Emotional State. IJIMAI. 5 (2019), 6, 33-42. https://doi.org/10.9781/ijimai.2019.03.004

[6] Chen, J., Hu, B., Moore, P., Zhang, X., & Ma, X. Electroencephalogram-based emotion assessment system using ontology and data mining techniques. Applied Soft Computing. 30 (2015), 663-674. https://doi.org/10.1016/j.asoc.2015.01.007.

[7] He, H., Tan, Y., Ying, J., & Zhang, W. Strengthen EEG-based emotion recognition using firefly integrated optimization algorithm. Applied Soft Computing. 94 (2020), 106426. https://doi.org/10.1016/j.asoc.2020.106426

[8] Zamkah, A., Hui, T., Andrews, S., Dey, N., Shi, F., & Sherratt, R. S. Identification of suitable biomarkers for stress and emotion detection for future personal affective wearable sensors. Biosensors. 10 (2020), 4, 40. https://doi.org/10.3390/bios10040040

[9] Ahmad F. Klaib, Nawaf O. Alsrehin, Wasen Y. Melhem, Haneen O. Bashtawi, Aws A. Magableh, Eye tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and Internet of Things technologies, Expert Systems with Applications. 166 (2021), 114037. https://doi.org/10.1016/j.eswa.2020.114037

[10] Alexandria M. Noble, Melissa Miles, Miguel A. Perez, Feng Guo, Sheila G. Klauer, Evaluating driver eye glance behavior and secondary task engagement while using driving automation systems, Accident Analysis & Prevention. 151 (2021),105959. https://doi.org/10.1016/j.aap.2020.105959.

[11] Bellamkonda, S., & Gopalan, N. P. An enhanced facial expression recognition model using local feature fusion of Gabor wavelets and local directionality patterns. International Journal of Ambient Computing and Intelligence. 11 (2020), 1, 48-70. https://doi.org/10.4018/IJACI.2020010103.

[12] Chakraborty, B., Bhattacharyya, S., & Chakraborty, S., Generative model based video shot boundary detection for automated surveillance, International Journal of Ambient Computing and Intelligence. 9(2018), 4, 69-95. https://doi.org/10.4018/IJACI.2018100105.

[13] Fischer, G. User Modeling in Human–Computer Interaction, User Model. User-Adapt. Interact. 11 (2001), 65–86. https://doi.org/10.1023/A:1011145532042.

[14] Cowie, R.; Douglas-Cowie, E.; Tsapatsoulis, N.; Votsis, G.; Kollias, S.; Fellenz, W.; Taylor, J., Emotion recognition in human-computer interaction, IEEE Signal Process. Mag. 18 (2001), 32–80. https://doi.org/ 10.1109/79.911197.

[15] Michael D. McNeese, New visions of human–computer interaction: making affect compute, International Journal of Human-Computer Studies. 59 (2003), 1-2, 33-53. https://doi.org/10.1016/S1071-5819(03)00059-4.

[16] Md Rakibul Mowla, Rachael I. Cano, Katie J. Dhuyvetter, David E. Thompson, Affective brain-computer interfaces: Choosing a meaningful performance measuring metric, Computers in Biology and Medicine. 126 (2020), 104001. https://doi.org/10.1016/j.compbiomed.2020.104001.

[17] Xiaohua Huang, Jukka Kortelainen, Guoying Zhao, Xiaobai Li, Antti Moilanen, Tapio Seppänen, Matti Pietikäinen, Multi-modal emotion analysis from facial expressions and electroencephalogram, Computer Vision and Image Understanding. 147 (2016), 114-124. https://doi.org/10.1016/j.cviu.2015.09.015.

[18] Jian Zhou, Shujie Chu, Xin Li, Fu Xiao, Lijuan Sun, An EEG emotion recognition method based on transfer learning and echo state network for HilCPS, Microprocessors and Microsystems. 2020, 103381, https://doi.org/10.1016/j.micpro.2020.103381.

[19] Hess EH, Polt JM, Pupil size as related to interest value of visual stimuli. Science. 132 (1960), 3423, 349-50. https://doi.org/10.1126/science.132.3423.349.

[20] Margaret M. Bradley, Laura Miccoli, Miguel A. Escrig, and Peter J. Lang, The pupil as a measure of emotional arousal and autonomic activation, Psychophysiology. 45 (2008), 4, 602–607. https://doi.org/10.1111/j.1469-8986.2008.00654.x.

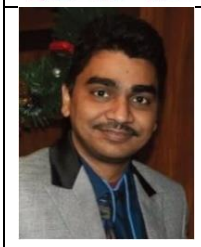
[21] Annika K., Helen N., Sarah S., Sinha E., Jan S., Deborah W., Beate M., Sebastian B., Sascha T., Anke W., Gerd W., Heinrich R., Christine K., Attentional bias in veterans with deployment-related posttraumatic stress disorder

before and after internet-based cognitive behavioral therapy - An eye-tracking investigation, Journal of Behavioral and Cognitive Therapy. 30 (2020), 4, 267-281. https://doi.org/10.1016/j.jbct.2020.03.003.
[22] Gregor Wilbertz, Madhura Ketkar, Matthias Guggenmos, Philipp Sterzer, Combined fMRI- and eye movement-based decoding of bistable plaid motion perception, NeuroImage, 171 (2018), 190-198. https://doi.org/10.1016/j.neuroimage.2017.12.094.
[23] C. Yen, M.C. Chiang, Examining the effect of online advertisement cues on human responses using eye-tracking, EEG, and MRI, Behavioural Brain Research 2021; 113128, https://doi.org/10.1016/j.bbr.2021.113128
[24] Y. Dong, F. Ren, Multi-reservoirs EEG signal feature sensing and recognition method based on generative adversarial networks, Computer Communications. 164 (2020), 177-184. https://doi.org/10.1016/j.comcom.2020.10.004.
[25] Y. Yao, J. Plested, T. Gedeon, Information-preserving feature filter for short-term EEG signals, Neurocomputing. 408 (2020), 91-99. https://doi.org/10.1016/j.neucom.2019.11.106.
[26] T. Tuncer, S. Dogan, F. Ertam, A. Subasi, A novel ensemble local graph structure based feature extraction network for EEG signal analysis, Biomedical Signal Processing and Control 61 (2020), 102006. https://doi.org/10.1016/j.bspc.2020.102006.
[27] T. Tuncer, A new stable nonlinear textural feature extraction method based EEG signal classification method using substitution Box of the Hamsi hash function: Hamsi pattern, Applied Acoustics. 172 (2021), 107607. https://doi.org/10.1016/j.apacoust.2020.107607.
[28] Zhong Yin, Mengyuan Zhao, Yongxiong Wang, Jingdong Yang, Jianhua Zhang, Recognition of emotions using multimodal physiological signals and an ensemble deep learning model, Computer Methods and Programs in Biomedicine. 140 (2017), 93-110. https://doi.org/10.1016/j.cmpb.2016.12.005.
[29] Md Sirajus Salekin, Ghada Zamzmi, Dmitry Goldgof, Rangachar Kasturi, Thao Ho, Yu Sun, Multimodal spatio-temporal deep learning approach for neonatal postoperative pain assessment, Computers in Biology and Medicine. 129 (2021), 104150. 10.1016/j.compbiomed.2020.104150.
[30] Mohammad-Parsa Hosseini, Tuyen X. Tran, Dario Pompili, Kost Elisevich, Hamid Soltanian-Zadeh, Multimodal data analysis of epileptic EEG and rs-fMRI via deep learning and edge computing, Artificial Intelligence in Medicine. 104 (2020), 101813. https://doi.org/10.1016/j.artmed.2020.101813.
[31] Mingqi Lv, Wei Xu, Tieming Chen, A hybrid deep convolutional and recurrent neural network for complex activity recognition using multimodal sensors, Neurocomputing. 362 (2019), 33-40. https://doi.org/10.1016/j.neucom.2019.06.051.
[32] Ali, M. N. Y., Sarowar, M. G., Rahman, M. L., Chaki, J., Dey, N., & Tavares, J. M. R. Adam deep learning with SOM for human sentiment classification. International Journal of Ambient Computing and Intelligence. 10 (2019), 3, 92-116. 10.4018/IJACI.2019070106.
[33] Y. Kuang, Q. Wu, Y. Wang, N. Dey, F. Shi, R. G. Crespo, R. S. Sherratt, Simplified inverse filter tracked affective acoustic signals classification incorporating deep convolutional neural networks, Applied Soft Computing. 97 (2020), Part A, 106775. https://doi.org/10.1016/j.asoc.2020.106775.
[34] Lan, K., Wang, D. T., Fong, S., Liu, L. S., Wong, K. K., & Dey, N. A survey of data mining and deep learning in bioinformatics. Journal of medical systems. 42 (2018), 8, 1-20. https://doi.org/10.1007/s10916-018-1003-9
[35] Haque, M. R., Hafiz, R., Al Azad, A., Adnan, Y., Mishu, S. A., Khatun, A., & Uddin, M. S. Crime Detection and Criminal Recognition to Intervene in Interpersonal Violence Using Deep Convolutional Neural Network With Transfer Learning. International Journal of Ambient Computing and Intelligence. 12 (2021), 4, 1-14. https://doi.org/10.4018/IJACI.20211001.oa1.
[36] Shibendu Mahata, Rajib Kar, Durbadal Mandal, Optimal fractional-order highpass Butterworth magnitude characteristics realization using current-mode filter, AEU - International Journal of Electronics and Communications. 102 (2019), 78-89. https://doi.org/10.1016/j.aeue.2019.02.014.
[37] Emanuele Maiorana, Deep learning for EEG-based biometric recognition, Neurocomputing, 410 (2020), 374-386. https://doi.org/10.1016/j.neucom.2020.06.009.
[38] Jarosław Bilski, Leszek Rutkowski, Jacek Smoląg, Dacheng Tao, A novel method for speed training acceleration of recurrent neural networks, Information Sciences. 553 (2021), 266-279. https://doi.org/10.1016/j.ins.2020.10.025
[39] Long C. Nguyen, H. Nguyen-Xuan, Deep learning for computational structural optimization, ISA Transactions. 103 (2020),177-191. https://doi.org/10.1016/j.isatra.2020.03.033
[40] Shenghao Tang, Changqing Shen, Dong Wang, Shuang Li, Weiguo Huang, Zhongkui Zhu, Adaptive deep feature learning network with Nesterov momentum and its application to rotating machinery fault diagnosis, Neurocomputing. 305 (2018), 1-14. https://doi.org/10.1016/j.neucom.2018.04.048.
[41] Bradley, M.M., & Lang, P.J. Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida, 1999.

[42] Bradley, M. M., & Lang, P. J., Affective Norms for English Text (ANET): Affective ratings of text and instruction manual. (Tech. Rep. No. D-1). University of Florida, Gainesville, FL, 2007.
[43] Kousta ST, Vigliocco G, Vinson DP, Andrews M, Del Campo E. The representation of abstract words: why emotion matters. J Exp Psychol Gen. 140 (2011), 1,14-34. https://doi.org/10.1037/a0021446.
[44] Lang, P.J., Herring, D.R., Duncan, C., Richter, J., Sege, C.T., Weymar, M., Limberg, A., Hamm, A.O., & Bradley, M.M. The startle-evoked potential: Negative affect and severity of pathology in anxiety/mood disorders. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. 3 (2018), 626-634. https://doi.org/10.1016/j.bpsc.2017.07.006.
[45] Robert Plutchik, Herman Van Praag. The measurement of suicidality, aggressivity and impulsivity, Progress in Neuro-Psychopharmacology and Biological Psychiatry. 13 (1989),S1, S23-S34. https://doi.org/10.1016/0278-5846(89)90107-3.
[46] Delorme A & Makeig S., EEGLAB: an open-source toolbox for analysis of single-trial EEG dynamics, Journal of Neuroscience Methods. 134 (2014), 9-21. https://doi.org/10.1016/j.jneumeth.2003.10.009.
[47] Dimigen, O., Optimizing the ICA-based removal of ocular EEG artifacts from free viewing experiments. NeuroImage. 207 (2020), 116117. https://doi.org/10.1016/j.neuroimage.2019.116117
[48] Tianhu Zhang, Xueyi You, Improvement of the Training and Normalization Method of Artificial Neural Network in the Prediction of Indoor Environment, Procedia Engineering. 121 (2015), 1245-1251. https://doi.org/10.1016/j.proeng.2015.09.152
[49] Gholami, Amir; Kwon, Kiseok; Wu, Bichen; Tai, Zizheng; Yue, Xiangyu; Jin, Peter; Zhao, Sicheng; Keutzer, Kurt, SqueezeNext: Hardware-Aware Neural Network Design. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2018, pp.1719-171909. https://doi.org/10.1109/CVPRW.2018.00215
[50] C. Szegedy et al., Going deeper with convolutions, in Proc. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, 2015; pp. 1-9.
[51] K. He, X. Zhang, S. Ren and J. Sun, Deep Residual Learning for Image Recognition, in Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016; pp. 770-778.
[52] Gayard, Laurent, Darknet: Geopolitics and Uses. Hoboken, NJ: John Wiley & Sons. 2018; pp. 158
[53] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, Alexander A Alemi, Inception-v4, inception-ResNet and the impact of residual connections on learning, AAAI'17: Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence 2017, pp. 4278-4284, ASM USA.

Authors Bios

Qun Wu, is an Associate Professor of Human Factor at the Institute of Universal Design, Zhejiang Sci-Tech University, China. He received his Ph.D. in College of Computer Science and Technology from Zhejiang University, China, in 2008. He holds a B.E. degree in Industrial Design from Nanchang University, China, in 2001, and an M.E. degree in Mechanical Engineering from Shaanxi University of Science and Technology, China, in 2004. Dr. Qu published over 40 journal papers and conference proceedings; his research interests include machine learning, human factor, and product innovation design.



Nilanjan Dey was born in Kolkata, India, in 1984. He received his B.Tech. Degree in Information Technology from the West Bengal University of Technology in 2005, M. Tech. in Information Technology in 2011 from the same University, and Ph.D. in digital image processing in 2015 from Jadavpur University, India. In 2011, he was appointed as an Asst. Professor in the Department of Information Technology at JIS College of Engineering, Kalyani, India, followed by Bengal College of Engineering College, Durgapur, India, in 2014. He is now employed as an Associate Professor in the Department of Computer Science and Engineering JIS University, Kolkata, India. His research topic is signal processing, machine learning, and information security. Dr. Dey is an Associate Editor of IEEE Access and is currently the Editor-in-Chief of the International Journal of Ambient Computing and Intelligence (IJACI), and Series co-editor of Springer Tracts of Nature-Inspired Computing (STNIC).

Fuqian Shi graduated from the College of Computer Science and Technology, Zhejiang University, got his Ph.D. in Engineering and was a visiting Associate Professor at the Department of Industrial Engineering and Management System, University of Central Florida, the USA, from 2012 to 2014. He is a Senior Member of IEEE, Membership of ACM; and sever as over 40 committee board member of international conferences; Dr. Shi also serves as an Associate Editors of the International Journal of Ambient Computing and Intelligence (IJACI), International Journal of Rough Sets, and Data Analysis (IJRSDA), and special issue editor of fuzzy engineering and intelligent transportation in INFORMATION: An International Interdisciplinary Journal. He published over 100 journal papers and conference proceedings; his research interests include fuzzy inference systems, artificial neural networks, biomechanical engineering, and bioinformatics.

	Rubén González Crespo, Dr. Rubén González Crespo has a PhD in Computer Science Engineering. Currently, he is Vice-Chancellor of Academic Affairs and Faculty from UNIR and Global Director of Engineering Schools from PROEDUCA Group. He is an advisory board member for the Ministry of Education at Colombia and evaluator from the National Agency for Quality Evaluation and Accreditation of Spain (ANECA). He is a member of different committees at ISO Organization. Finally, He has published more than 200 papers in indexed journals and congresses.
	R. Simon Sherratt is currently a Professor of Biomedical Engineering at the University of Reading, UK. Professor Simon Sherratt received the B.Eng. From Sheffield City Polytechnic (now Sheffield Hallam University), M.Sc. from The University of Salford, and Ph.D. from The University of Salford; he was elected as Fellow of the IEEE in 2012, Fellow of the IET in 2009; Senior Fellow of the Higher Education Academy in 2014. He is a Chartered Engineer (C.Eng.) and registered European Engineer (Eur Ing). Professor Simon Sherratt was awarded the IEEE International Symposium on Consumer Electronics (ISCE) 2006 1st Place Best Paper Award: IEEE Chester Sall Award for best papers in the IEEE Transactions on Consumer Electronics in 2006, 2016, 2017, 2018. He has published over 200 articles in peer review journals and international conferences. His research area is wearable devices, mainly for healthcare and emotion detection.