

# Eye-Tracking Signals Based Affective Classification Employing Deep Gradient Convolutional Neural Networks

Yuanfeng Li<sup>1</sup>, Jiangang Deng<sup>1</sup>, Qun Wu<sup>2</sup>, Ying Wang<sup>3\*</sup>

<sup>1</sup> Jiyang College of Zhejiang A&F University, Hangzhou (PR China)

<sup>2</sup> Institute of Universal Design, Zhejiang Sci-Tech University, Hangzhou (PR China)

<sup>3</sup> Department of Industrial Design at College of Art and Design, Zhejiang Sci-Tech University, Hangzhou (PR China)

Received 9 December 2020 | Accepted 9 April 2021 | Published 17 June 2021



## ABSTRACT

Utilizing biomedical signals as a basis to calculate the human affective states is an essential issue of affective computing (AC). With the in-depth research on affective signals, the combination of multi-model cognition and physiological indicators, the establishment of a dynamic and complete database, and the addition of high-tech innovative products become recent trends in AC. This research aims to develop a deep gradient convolutional neural network (DGCNN) for classifying affection by using an eye-tracking signals. General signal process tools and pre-processing methods were applied firstly, such as Kalman filter, windowing with hamming, short-time Fourier transform (SIFT), and fast Fourier transform (FTT). Secondly, the eye-moving and tracking signals were converted into images. A convolutional neural networks-based training structure was subsequently applied; the experimental dataset was acquired by an eye-tracking device by assigning four affective stimuli (nervous, calm, happy, and sad) of 16 participants. Finally, the performance of DGCNN was compared with a decision tree (DT), Bayesian Gaussian model (BGM), and k-nearest neighbor (KNN) by using indices of true positive rate (TPR) and false negative rate (FPR). Customizing mini-batch, loss, learning rate, and gradients definition for the training structure of the deep neural network was also deployed finally. The predictive classification matrix showed the effectiveness of the proposed method for eye moving and tracking signals, which performs more than 87.2% in accuracy. This research provided a feasible way to find more natural human-computer interaction through eye moving and tracking signals and has potential application on the affective production design process.

## KEYWORDS

Affective Computing, Deep Gradient Convolutional Neural Networks, Eye Moving and Tracking Signal, Short Time Fourier Transform.

DOI: 10.9781/ijimai.2021.06.002

## I. INTRODUCTION

**H**UMAN affective comes from the physiological response of the machine to the periphery. While specific physiological activities generate different affective experiences, such as fear, may cause the accelerated heart rate, irregular breathing rhythm, abnormal skin electrical response, and the formation of a corresponding [1]-[3]. The brain mainly drives the dominant affective; under laboratory conditions, evoking an individual's real affective experience and acquiring physiological signals at the same time has been regarded as a popular affective measurement method currently. The affective generation method may be divided into material stimulation and situation induction [4]. Previous studies have shown that, under affective stimulation, obtaining human stress data through physiological signals and then analyzing it is a relatively stable and

reliable method [5], including skin temperature, electromyography (EMG), electrocardiogram (ECG), blood volume, etc. [6]. To data, affective recognition based on multiple physiological signals is more complicated; firstly, the processing of physiological signals is more complicated, and the denoising algorithm needs to be developed stably continuously [7]. Weak noise may cause significant overall changes finally. Secondly, various physiological signals have unique denoising methods; common noises include motion artifacts, power frequency noise, and baseline drift caused by sensor movement. Besides, the qualitative problem of affective is currently more complicated, and it is mainly divided into discrete affective definitions and dimensional definitions. Either definition requires a self-assessment or manual labeling process, which brings a lot of subjectivity to the overall framework of affective recognition in this research. Now psychology does not conclude that the method is good, but from an experimental point of view, discrete affective are better recognized when the types of affective signal are few and clear. Finally, the physiological signals come from personal differences. In some scenes, there may be people who reflect strongly that some people have no fluctuations. The physiological signal is also that a person may have blood pressure

\* Corresponding author.

E-mail address: wangying1980@163.com

and heartbeat that is already fast. There are two ways to solve this state. Among them, detecting the primary form in a calm state and regularizing the features in the affective state is acceptable [8].

This paper studies the recognition of human affective by eye movement signals, uses efficient preprocessing methods and feature calculations, and uses deep neural networks to complete the precise classification of affective. Eye movement signal extraction methods include human eye pupil positioning and eye movement feature extraction; among them, techniques such as pupil center positioning based on gray-scale information are performed. The eye movement signal is a signal of potential changes around the eye generated by eye movement. The signal has the advantages of high amplitude, easy waveform recognition, and simple processing. At present, eye-tracking signals for human-computer interaction research are a hot issue and a relatively new direction [9]-[11]. Some scholars studied people's mental activities by examining their eye movements and explored the relationship between eye movements and human mental activities by analyzing the recorded eye movement data. The eye tracker's advent provides a new way for psychologists to use eye moving and tracking techniques to explore the visual information processing mechanism of humans under various conditions and observe their wonderful or interesting relationship directly or indirectly with mental activities [12] [13]. The eye movement technology has experienced observation methods, post-image methods, mechanical recording methods, optical recording methods, image recording methods, and other methods.

Eye moving and tracking technology extracts data features such as fixation point, fixation time and frequency, saccade distance, pupil size, etc. From the recording of eye movement trajectory, to study an individual's internal cognitive process, Juhola et al. used eye movement signal analysis for otoneurologic patients [14][15]; Kasneci, E., et al. employed eye-tracking and aggregated physiological signals for perception prediction [16]. Eye-tracking also may be applied for computer-aided diagnosis [17], equipment operator [18], and video learning [19][20]. The eye-tracking signals were also successfully applied to detect autism patients [21]-[23]. There are three basic ways of eye movement: fixation, saccades, and pursuit movement. Eye movement reflects the selection mode of visual information, which greatly impacts discovering the psychological mechanism of cognitive processing. From the research report, the commonly used data or parameters of psychological research using eye trackers mainly include gaze point trajectory map, eye movement time, average velocity, amplitude time and distance, pupil size (area or diameter, unit pixel) and blink. The spatiotemporal feature of eye movement is the physiological and behavioral performance in visual information extraction. It has a direct or indirect relationship with human psychological activities, so many psychologists are devoted to eye movement research. When the eyeball moves, a weak magnetic field is formed between the retina and the cornea. According to the study, the potential change between the cornea and the retina is highly correlated with the eyeball's rotation angle. It is linearly correlated between 0-30, and 30-60 is the relationship between the sine and cosine line. Electrodes may not be directly assigned to the cornea for measurement while skin electrical tests are performed on the eye's outer skin. But this is different from surface EMG (sEMG), for the test range and signal are weak and stable information cannot be obtained. Therefore, the currently effective method is to set a mirror on the cornea or iris for optical measurement. Among them, the contact eyepiece is reflected light. The other method is the detection coil method, which determines the direction of the eye movement by the change of the magnetic field around the eye. At present, electro-oculography (EOG) signal method is also popular [24].

A deep neural network was applied for classifying eye movement signals recently. The simplest artificial neural network is a binary linear

classifier; a neuron's structure can be divided into dendrites, synapses, cell bodies, and axons. A single neuron can be regarded as a machine with only two states. The transformation of a neuron depends on the number of input signals received from other neurons and the synapses' strength. When the semaphores' sum exceeds a certain threshold, the neuron body will be excited and generate electrical pulses [25][26]. Electrical pulses are transmitted along the axon and through the synapse to other neurons which is defined as a synapse, bias as a threshold, and activation function as neuron body. The deep network of unsupervised learning is aimed at pattern analysis or synthesis tasks, capturing high-order correlations of observed or visible data without target label information; and the supervised learning deep network directly provides the discriminative ability for pattern classification. Describes the posterior distribution under visible data, also called discriminative deep network (DDN). Besides, there is another type called hybrid deep network (HDN) based on a discriminative model. The unsupervised deep network mainly includes deep Boltzmann machine, sum-product networks (SPN), recurrent neural network (RNN), etc. The training process of deep neural networks is critical in the process of signal classification. How to make the network perform well on the training set and make the network perform the same on the test set if the training set performs well is critical. How to tune on the training set is also a vital topic, including how to choose the appropriate loss function, mini-batch, choose a new activation function, adaptive learning rate, and momentum. The current practice is to increase the training set, stop early, regularize, dropout, and improve network structure [27]. The deep convolutional neural network (DCNN) has developed rapidly and was initially used for image recognition classification [28]. Using the convolution layer and the pooling layer, the ability to accurately predict the image is accelerated. RNN and its derived algorithms can be used for speech recognition, natural language processing, speech synthesis, etc. The deep neural network (DNN) can model changes in time series. The research of DNN for classification is relatively mature, such as LeNet, AlexNet, ZFNet, VGGNet, GoogLeNet, Inception-v1, etc. [29]-[31]. Applied eye movement signals for affective classification is feasible due to the deep learning technologies applied in other datasets, such as images [32] and wide applications in industrial engineering [33] [34].

This paper developed a novel deep neural network based on gradients calculation by converting eye moving and tracking signals to images; feature extraction and evaluation indices were also defined for the comparing analysis and finally performed the proposed model for the dataset. The organization of the remaining sections is as follows. Section II introduces modeling for the dataset, including preprocessing algorithms. Section III addresses the results and comparing analysis. Section IV involves the concluding remarks of the studies and future works.

## II. MODELING

### A. Preprocessing

The specific requirements for signal preprocessing are due to the features of the vibration signal itself. The function of signal preprocessing is to make a certain extent by using important factors for influencing subsequent signal analysis. The fast Fourier transform (FFT) is a general term for efficient and fast calculation methods through calculating discrete Fourier transform (DFT). DFT can discretize the finite-length sequence in the frequency domain, but its calculation is too large to handle the problem in real-time; the basic idea of FFT is to sequentially decompose the original N-point sequence into a series of short sequences [35][36]. Make full use of the symmetric and periodic nature of the exponential factors in the DFT calculation formula, and then find the corresponding DFT of these short sequences and make appropriate combinations to achieve the purpose of eliminating duplicate calculations, reducing multiplication

operations, and simplifying the structure. Fast algorithms such as high basis and split basis have been developed, such as the Winograd Fourier transform algorithm (WFTA) [37] and prime factors based on number theory and polynomial theory Fourier transform algorithm. Their common feature is that when  $N$  is a prime number, the DFT calculation can be converted into a circular convolution, reducing the number of multiplications and increasing the speed [38]. The obvious advantage of a small calculation amount makes FFT widely used in signal processing technology, and real-time processing of signals can be realized in combination with high-speed hardware.

Another preprocessing method is signal filtering (i.e., wave filtering), an operation to filter out specific frequency bands in the signal and is an important measure to suppress and prevent interference. Filtering is divided into classic filtering and modern filtering. Classical filtering is an engineering concept based on Fourier analysis and transformation. According to higher mathematics theory, any signal that satisfies certain conditions can be regarded as a superposition of infinite sine waves. In other words, the signal is a linear superposition of sine waves of different frequencies. The sine waves of different frequencies. The sine waves of different frequencies that make up the signal are called the signal's frequency component or the harmonic component. The Kalman filter used in this study is a recursive algorithm (i.e., real-time algorithm). Specifically, for discrete-time filtering, as long as the dimension of  $X$  is appropriately increased, the filter value table at time  $t$  can be some linear combination of the filter value at the previous time and the observation value  $Y(t)$  at the current time. For continuous-time filtering, the linear stochastic differential equations that should be satisfied with  $Y(t)$  may be given. In the case where the observation results and output filter values need to be continuously increased, such an algorithm speeds up the processing of data and reduces the amount of data storage.

First, the following filter equations need to be used for determining parameters here. The Kalman filter is suitable for linear systems, and the state equation and observation equation of the system are,

$$x(k) = Ax(k-1) + Bu(k) + w(k) \quad (1)$$

$$z(k) = Hx(k) + y(k) \quad (2)$$

Where,  $x(k)$  is the state of the system at time  $k$ ;  $u(k)$  is the control quantity of  $x(k)$ ;  $w(k)$  is to process the noise conforming to the Gaussian distribution; and the is that,  $z(k)$  is the observation value of the system at time  $k$  while  $y(k)$  measures the noise conforming to the Gaussian distribution. The covariance is  $A$ ,  $B$  and  $H$  present the system parameters, matrix for multiple input and multiple output, and several constants for single input and single output separately. The algorithm 1 describes the filter in detail [39].

---

#### Algorithm 1: Kalman Filter

**Require:** data: input signal; Q: covariance of  $w(k)$ ; R: covariance of  $z(k)$ ;  $x_0$ ;  $p_0$ .

**Output:** P

$L \leftarrow \text{length}(\text{data})$  # calculate the length of the input data

$K \leftarrow \text{zero}(L,1)$  # set the initial zeros to the parameters

$X \leftarrow \text{zero}(L,1)$

$P \leftarrow \text{zero}(L,1)$

$X(1) \leftarrow x_0$ ; # set the start point

$P(1) \leftarrow p_0$

**FOR**  $i$  in (2 to  $L$ )

$K(i) \leftarrow P(i-1) / (P(i-1) + R)$

$X(i) \leftarrow X(i-1) + K(i) * (\text{data}(i) - X(i-1))$

$P(i) \leftarrow P(i-1) - K(i) * P(i-1) + Q$

**ENDFOR**

---

Another preprocessing algorithm is the short-time Fourier transform (STFT), a variant of the Fourier transforms, also known as windowed Fourier transform or time-dependent Fourier transform, used to determine the sinusoidal frequency and phase of a local portion of a signal that changes with time. The process of calculating the STFT is to divide the long-term signal into several shorter equal-length signals and then calculate the Fourier transform of each shorter segment separately. It **usually describes** changes in the frequency and time domains and as one of the **essential** tools in time-frequency analysis [40]. Continuously, we have that,

$$X(f) = \int_{-\infty}^{\infty} \frac{x(t)}{e^{2\pi ift}} dt \quad (3)$$

$$X(t, f) = \int_{-\infty}^{\infty} \frac{\varpi(t - \tau)x(\tau)}{e^{2\pi if\tau}} d\tau \quad (4)$$

Where,  $f$  is frequency of the signal;  $t$  is time. The STFT is to truncate the original Fourier transform into multiple segments in the time domain and perform the Fourier transform separately. Each segment is recorded as time  $t_i$ , and the frequency domain characteristics are obtained by corresponding FFT, and the time  $t_i$  can be roughly estimated. The frequency domain characteristics of time (that is, the corresponding relationship between the time domain and the frequency domain is simultaneously guided). The tool used for signal truncation is called a window function (the width is equivalent to the length of time). The smaller the window, the more obvious the time-domain characteristics, but currently, for the number of points is too small, the FFT reduces the accuracy and the frequency-domain characteristics are not obvious. In order to ensure the improvement of the time domain characteristics on the basis of the frequency domain characteristics, it is often selected to overlap a part of the front and rear window functions, so that the time of the two windows is determined is closer to improve the time domain analysis ability. However, it is not better to have more overlaps. Too many overlapping points will greatly increase the amount of calculation which is resulting in low efficiency. Therefore, the number of overlapping points in the front and rear windows also needs to be determined. Several windowed models are introduced as below,

- Hamming window:

$$w(n) = 0.54 - 0.46\cos\left(\frac{n}{N}\right) \quad (5)$$

- Hanning window:

$$w(n) = \frac{\left(1 - \cos\left(\frac{n}{N}\right)\right)}{2} \quad (6)$$

- Rectangular window

$$w(n) = 1.0 \quad (7)$$

- Triangle window

$$w(n) = TRI\left(\frac{2n}{N}\right) \quad (8)$$

- Blackman, third order raised cosine window

$$w(n) = 0.42 + 0.5\cos\left(\frac{n}{N}\right) + 0.08\cos\left(\frac{2n}{N}\right) \quad (9)$$

- Blackman-Harris window

$$w(n) = 0.359 - 0.489\cos\left(\frac{n}{N}\right) + 0.141\cos\left(\frac{2n}{N}\right) - 0.012\cos\left(\frac{3n}{N}\right) \quad (10)$$

In this research, the window function selects the hamming window, and the maximum number of DFT points is not greater than 256; user

input (pass value) is signal, window, overlap, N, fs, etc. According to the size of the window, split the signal and multiply it with the window function; perform N-point FFT on each signal segment and find the energy spectral density; Algorithm 2 described the windowing process on eye movement signals, and splitting the signals for further analysis was described in Algorithm 3 [41].

**Algorithm 2** preprocessing the eye movement signal by windowing

**Required:** signal.vector: original signal sequence; N: length of signal; w: window length; noverlap: the number of overlapping windows; nfft: FFT/DFT points; fs: sampling frequency;

**Output:** signal

```

FOR i in (0 to w) #hamming window
    hamming[i] = 0.54 - 0.46 * cos(2*M_PI*i/(w-1))
    windowPV ← windowPV+ pow(hammingW[i], 2)
ENDFOR
#calculate the number of rows and columns of the short-time
Fourier transform signal array, the number of rows is the number of
time points, the number of columns is w
row ← (N - noverlap)/ (w - noverlap)
column ← w
half_Nfft ← nfft/2+1
FOR I in (0 to row)
    timeV[i] ← ((float)i)/((float)(fs*(w/2+1+(w-noverlap)*i)))
ENFOR
# separating the signals by row and column using windowing
FOR i in (0 to row)
    FOR j in (0 to column)
        signalXY[i][j] ← signal.vector[i*(w - noverlap) + j]
        signalXY[i][j] ← signalXY[i][j] *hammingW[j]
    ENDFOR
ENDFOR

```

## B. Feature Extraction Methods

The feature extraction (FE) methods include time domain, frequency domain and model-based method. For time-domain based features, there are waveform, pulse, kurtosis, margin, peak and zero crossing rate (ZCR), etc. the ZCR is calculated by,

$$ZCR_n = \frac{1}{2} \sum_{i=0}^{N-1} |\text{sgn}[x_n(i)] - \text{sgn}[x_n(i-1)]| \quad (11)$$

Where, N is the length of the frame, n is the number of frames;

$$\text{sgn}(x) = \begin{cases} 1 & x > 0 \\ -1 & x < 0. \end{cases}$$

Using time as an independent variable to describe changes in physical quantities is the most basic and intuitive expression of signals. In the time domain, the signal is filtered, amplified, statistical feature calculation, correlation analysis and other processing, collectively called the time domain analysis of the signal. Different eye movement waveforms have obvious differences. It is also feasible to directly analyze the waveform characteristics, such as amplitude and wavelength. In addition, the digital signal can also be fitted by the model method. As commonly used linear predictive coding (LPC). For eye movement signal, frequency domain-based feature methods mainly focus on power spectrum density (PSD) estimation. Here, we adopted parametric method, supposed that., eye movement signal is  $x(n)$ , the autocorrelation function of  $x(n)$  is  $r(k)$ , so, the PSD is,

$$p(w) = \sum_{k=-\infty}^{+\infty} r(k)e^{-jwk} \quad (12)$$

**Algorithm 3:** FFT and STFT preprocessing for eye moving and tracking signals

**Required:** nfft: FFT/DFT points; fs: sampling frequency

**Output:** S; P;

```

freqStep ← fs/nfft; freq[0] ← 0.0
FOR i in (1 to nfft)
    freq[i] ← freq_Step + freq[i-1]
ENDFOR
FOR i in (0 to row)
    FOR j in (0 to half_Nfft)
        Sxx[i][j] = (fft_Real[j]*fft_Real[j] + fft_Img[j]*fft_Img[j])/
        windowPV
    ENDFOR
ENDFOR
Pxx[0][0] ← Sxx[0][0]/fs
log_Pxx[0][0] ← 10*log10f(fabsf(Pxx[0][0]))
pxx_Max ← log_Pxx[0][0]
pxx_Min ← log_Pxx[0][0]
FOR i in (1 to row)
    Pxx[i][0] ← Sxx[i][0]/fsFloat
    log_Pxx[i][0] ← 10*log10f(fabsf(Pxx[i][0]))
    IF log_Pxx[i][0] > pxx_Max THEN
        pxx_Max ← log_Pxx[i][0]
    ENDIF
    IF log_Pxx[i][0] < pxx_Min THEN
        pxx_Min ← log_Pxx[i][0]
    ENDIF
    IF nfft mod 2 = 0 THEN
        FOR i in (0 to row)
            FOR j in (1 to half_Nfft)
                Pxx[i][j] ← Sxx[i][j] *2.0/fs
                log_Pxx[i][j] ← 10*log10(abs (Pxx[i][j])) #
                abs():absolute
                IF log_Pxx[i][j] > pxx_Max THEN
                    pxx_Max ← log_Pxx[i][j]
                ENDIF
                IF log_Pxx[i][j] < pxx_Min THEN
                    pxx_Min ← log_Pxx[i][j]
                ENDIF
            ENDFOR
        ENDFOR
    ELSE
        FOR i in (0 to row)
            Pxx[i][half_Nfft-1] ← Sxx[i][half_Nfft-1]/((float)fs)
            log_Pxx[i][half_Nfft-1] ← 10*log10(abs (Pxx[i]
            [half_Nfft-1]))
            IF log_Pxx[i][half_Nfft-1] > pxx_Max THEN
                pxx_Max ← log_Pxx[i][half_Nfft-1]
            ENDIF
        ENDFOR
        FOR i in (0 to row)
            FOR j in (1 to halfNfft-1)
                Pxx[i][j] ← Sxx[i][j] *2.0/((float)fs)
                log_Pxx[i][j] ← 10*log10f(fabsf(Pxx[i][j]))
                IF log_Pxx[i][j] > pxx_Max THEN
                    pxx_Max ← log_Pxx[i][j]
                ENDIF
                IF log_Pxx[i][j] < pxx_Min
                    pxx_Min ← log_Pxx[i][j]
                ENDIF
            ENDFOR
        ENDFOR
    ENDFOR
ENDFOR

```



Where,  $r(k) = EP[x(n)x^*(n+k)]^*$  is conjugate; EP is expectation.

If the length of eye movement signal is smaller, the PSD may be adjusted by,

$$\hat{P}(w) = \sum_{k=-(N-1)}^N r(k)e^{-jwk} \quad (13)$$

$$\text{Where, } \hat{r}(k) = \frac{1}{N-k} \sum_{n=k+1}^N x(n)x^*(n-k), 0 \leq k \leq N-1.$$

Parameter spectral density estimation first uses a certain function to fit the signal, then you can determine each parameter of the function model, and finally get the spectral characteristics of the parameter. When the signal fitted by the established function model is like the actual signal, the efficiency of the parameter spectral density will be relatively high. The commonly used models of modern spectral density are automatic moving average (ARMA), auto regressive (AR), moving average (MA), etc. For eye movement models, AR models are often used. The main advantage is that under high SNR conditions, a relatively high resolution can be obtained. It is suitable for short data processing. AR is a linear regression model that linearly combines random variables at several times in the early period to describe random variables at a certain period in the later period. It is essentially a linear prediction, expressed as follows:

$$x(n) = -\sum_{i=1}^c a_c(i)x(n-i) + \varepsilon(n) \quad (14)$$

Where,  $\varepsilon(n)$  is a white noise sequence with variance  $\sigma^2$  and mean is zero, and  $c$  represents the order of the AR model. Therefore, the eye movement signal sequence  $x(n)$  may be regarded as the output of the white noise sequence  $\varepsilon(n)$  through the AR model. When building an AR model, the first question is to determine the appropriate order. The order of the model is implemented in the recursion process. When using the LevinsonDurbin recursion method (LRM), each set of parameters from low order to high order can be given. When the minimum prediction error power of the model no longer changes, the correct order is acquired.

In this research, frequency domain indicators are also applied including center of gravity frequency (CGF), mean square frequency (MSF), root mean square frequency (RMSF), frequency variance (FV), frequency standard deviation (FSD), short-term power spectral density (STPSD), spectral entropy (SE), fundamental frequency (FF), and formant (F). We combined all signal features and used transform processing technologies to prepare the inputs for deep neural networks. The candidate transforms include FFT, non-parametric power spectrum (i.e., periodic graph method, Welch method), parametric power spectrum estimation method, STFT, wavelet transform, Hilbert transform, MFCC, Wigner distribution (WDF), Radon transform, and Gabor transform.

### C. Deep Gradients Convolutional Neural Network-based Classification Model

The gradient is a concept related to the directional derivative. The direction of the gradient is given by the angle of the gradient vector relative to the x-axis. The original gradient descent algorithm is obtained from the directional derivative, and then the momentum gradient descent algorithm is described. Due to the importance of hyperparameter learning rate to gradient descent, the gradient algorithm has multiple adaptive gradient descent algorithms; the forms of gradient descent include batch gradient descent (BDG), stochastic gradient descent (SGD), and mini-batch gradient descent (MGD); the evolution of gradient descent is mainly several adaptive gradient descent algorithms such as AdaGrad, RMSprop, AdaDelta,

**Algorithm 4** deep gradients convolutional neural network for eye movement signals classification

**Required:** numObservations: number of observations; miniBatchSize; maxEpochs; numIterationsPerEpoch

**Output:** loss; learnRate; epoch; Elapsed time

Iteration  $\leftarrow$  0

numIterationsPerEpoch  $\leftarrow$  floor(numObservations./miniBatchSize);

start  $\leftarrow$  tic

**FOR** epoch in (1 to maxEpochs)

idx  $\leftarrow$  randperm(numObservations)

XTrain  $\leftarrow$  XTrain(idx)

YTrain  $\leftarrow$  YTrain(idx)

**FOR** i in (1: numIterationsPerEpoch)

iteration  $\leftarrow$  iteration + 1

#Read mini-batch of data and apply the transformSequences

idx  $\leftarrow$  (i-1)\*miniBatchSize+1:i\*miniBatchSize;

[X,Y,numTimeSteps]  $\leftarrow$

transformSequences(XTrain(idx),YTrain(idx));

dIX  $\leftarrow$  darray(X)

# Evaluate the model gradients and loss using dlfeval.

[gradients, loss]  $\leftarrow$

dlfeval(@modelGradients,

dIX,Y,parameters,hyperparameters,numTimeSteps)

#Clip the gradients.

gradients  $\leftarrow$  dupdate(@g)

thresholdL2Norm(g.gradientThreshold,gradients);

# Adam optimizer for updating the network

[parameters,trailingAvg,trailingAvgSq]  $\leftarrow$

adamupdate(parameters,gradients, ...

trailingAvg, trailingAvgSq, iteration, learnRate)

**IF** plots == "training-progress"

D  $\leftarrow$  duration (0,0, toc(start),'Format','hh:mm:ss')

loss  $\leftarrow$  mean (loss/ numTimeSteps)

loss  $\leftarrow$  double(gather(extractdata(loss)))

loss  $\leftarrow$  mean(loss)

addpoints(lineLossTrain,iteration, mean(loss));

title ("Epoch: " + epoch + ", Elapsed: " + string(D))

drawnow # plot the results

**ENDIF**

**ENDFOR**

**IF** mod(epoch,learnRateDropPeriod) == 0

learnRate = learnRate\*learnRateDropFactor;

**ENDIF**

**ENDFOR**

and Adam. Gradient descent is a commonly used optimization method for machine learning. The difficulty of gradient descent is mainly reflected in the setting of the learning rate.

The minimum point, the first difficulty in saddle point gradient descent optimization, is the problem of setting the learning rate mentioned above. When the learning rate is too low, the convergence speed is slow, and when the learning rate is too large, it will cause training shocks and may diverge. In contrast, non-convex error functions generally appear in neural networks. When optimizing such functions, another difficulty is that the gradient descent process may fall into a local minimum. Studies have also pointed out that this

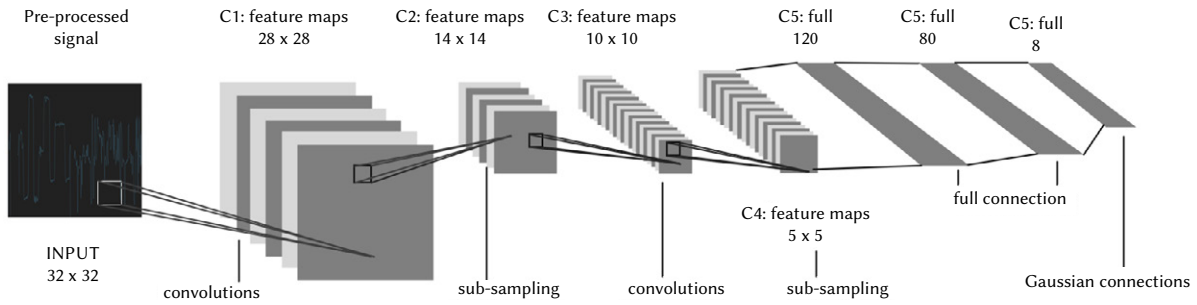


Fig. 1. deep gradients convolutional neural network for eye movement signals classification.

difficulty does not come from the local minimum, but more from the saddle point, those that are increasing in one dimension and decreasing in another dimension. Points with the same error usually surround these saddle points. Because the gradient in any size is approximately 0, it is difficult for SGD to escape from these saddle points.

The neural network's main task is to find the optimal parameters (weights and biases) during learning. This optimal parameter is also the parameter when the loss function is minimal. However, in general, the loss function is more complicated, and there are many parameters, so it is impossible to determine where to obtain the minimum value. So, the gradient method is the method to find the minimum value (or as small as possible) through the gradient. It is noted that the gradient indicates the direction in which the function value at each point decreases the most, so the path of the gradient does not necessarily point to the minimum. But along its direction can minimize the value of the function. Therefore, when looking for the minimum value (or as small as possible) of the function, use the information of the gradient as a clue to determine the direction of progress. Currently, the gradient method comes in handy. In the gradient method, the value of the function advances a certain distance from the current position along the gradient direction, then recalculates the gradient in the new direction, and then advances along the new gradient direction, and so on. Like this, the process of gradually decreasing the value of the function by continuously advancing in the direction of the gradient is the gradient method (gradient method). In neural networks (deep learning), the gradient method mainly refers to the gradient descent method [39]. We designed a deep gradients convolutional neural network-based classification model, which is illustrated in Fig.1.

Algorithm 4 introduced the process of the network proposed.

### III. RESULTS AND DISCUSSION

#### A. Data Acquisition and Preprocessing

Normally, the amplitude of the eye movement signal is in the range of 0.4-10mv, and the frequency is between 0-38Hz; the main frequency is 0-10Hz. While in this experiment, the Sampling frequency used for eye tracking is 600Hz. Eye image stream frequency is approximately 10 Hz. Accuracy is 0.3° at optimal conditions (down to 0.16°); 8-bit timestamped data (256 event codes) event-driven detection with a timestamp accuracy of 50 μs; Operating distance is of 55 to 75 cm from the eye tracker reference point; freedom of head movement is 34 cm width x 26 cm height at 65 cm, that means at least one eye tracked. Two cameras capture stereo images of both eyes for accurate measurement of eye gaze and position in space. The experiment uses a "non-invasive" technology based on video oculographic (VOG). The specific model is Tobii's X6, 0/120 series of bare machines; the user's range of activities is limited to a square of 2 meters, and authorized test personnel are required to place their chin on a fixed bracket to obtain more accurate data. The system compensates the head movement, and the specified head movement range is 44×22×30cm (length, width

and height). This experiment collected 16 group signals (8 males, 8 females, 18-22 years old). Fig. 2 shows the original tracking patterns, in which the black dots are the location, lines between dots are the traces. Fig. 3 shows the signals processed by using SIFT. FFT applied for eye-tracking signals is acquired from the experimental platform. (shown in Fig. 4)

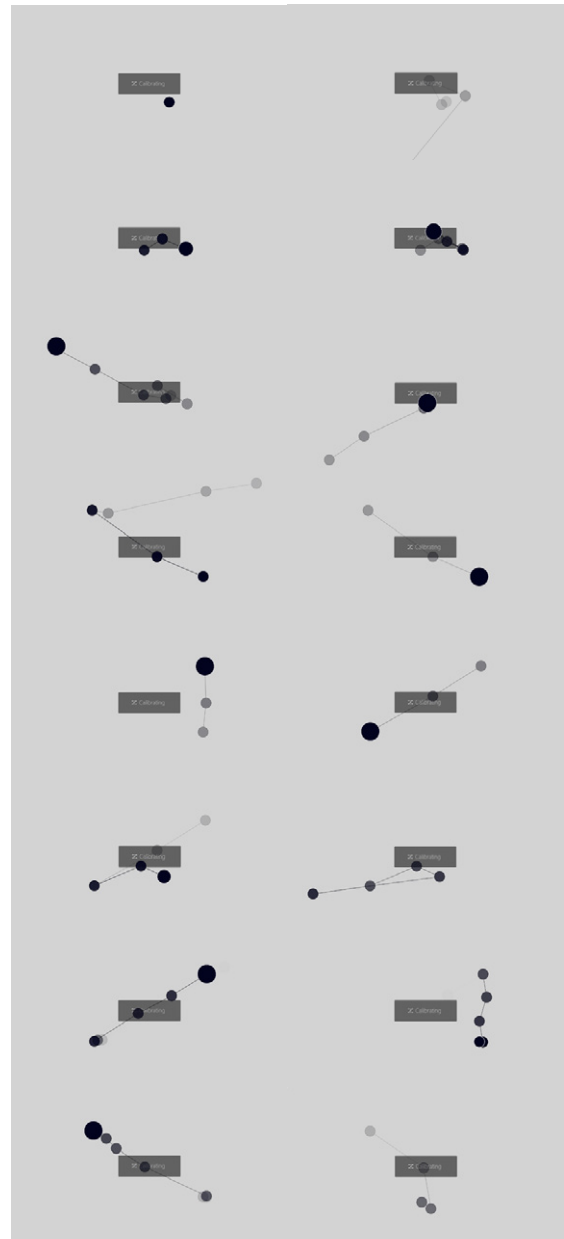


Fig. 2. Eye movement tracking scene experiment.

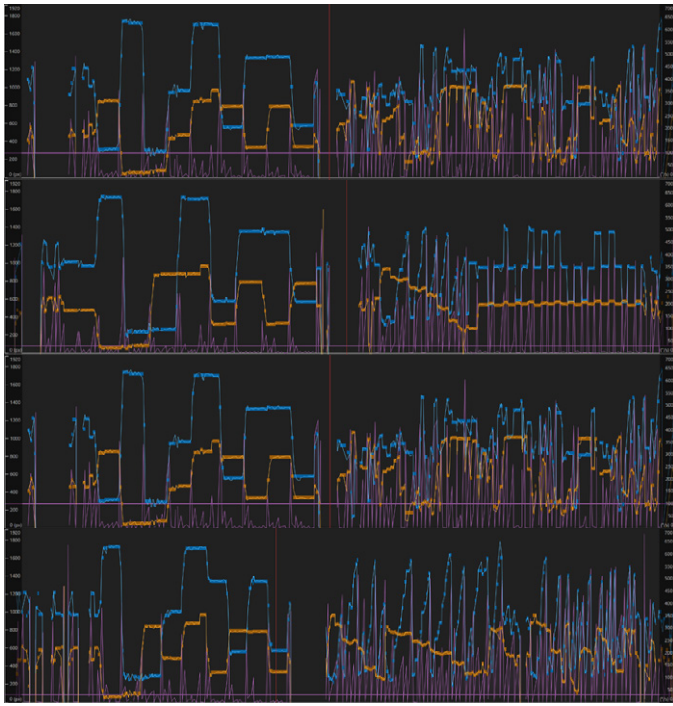


Fig. 3. Eye movement preprocessed signals acquired by eye tracking lab system Tobii pro.

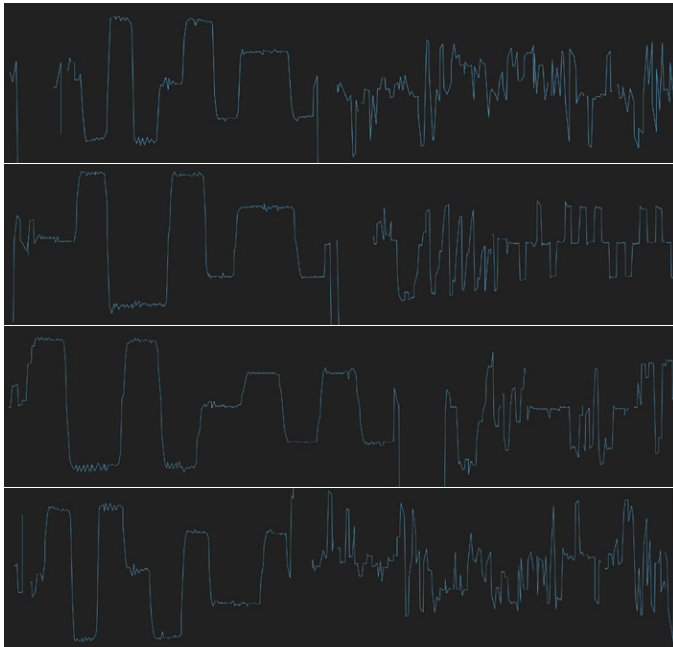


Fig. 4. FFT applied for Eye movement signals.

### B. Classification Results and Discussion

Frequency, sample frequency, eye gaze, eye position, moving speed, peak value of signals, time-synchronized average (TSA), amplify, root mean square, Impulse Factor, Signal-to-Noise Ratio (SNR), Total Harmonic Distortion (THD). The input features dataset is listed in Appendix A. The original training dataset for eye movement signals with features and manual labels are illustrated in Fig. 5. The features were labeled by a group of person in advance using investigation system. We have the 600 rows with 13 columns which including 12 features and 1 label for 4 affective of nervous, calm, happy, and sad. We also developed an evaluation index table for the performance of the

classification model. TP is true positive, TN is true negative, FP is false positive, and FN is false negative. Then we have that, Accuracy (A) is for the entire model, the number of prediction pairs is higher than the total number of samples. Accuracy is  $TP+TN/(TP+TN+FP+FN)$ . Precision (P) is the prediction is correct for a certain target, which is compared to the total number of that type predicted, P is  $TP/(TP+FP)$ . Recall (R) or recall rate is for a certain category that how much of the category has been predicted in total. R is  $TP/(TP+FN)$ . F1-score is for the case of the precision rate and the recall rate are always opposites; F1-score is an indicator that can balance the two to choose an optimal value. F1 is  $2(PR)/(P+R)$ . The proposed DGCNN classification results are shown in Fig. 6, which is the accuracy predictive matrix and AOC for DGCNN classification model and Fig. 7 shows the ROC curvature for the TPR and FNR by selecting affective of "calm"; the average accuracy of effective is more than 87.2%. Fig. 8 shows the predictive class matrix of decision tree (DT) model for eye-tracking signals; Fig. 9 shows the predictive class matrix of Gaussian naïve Bayes, and Fig. 10 shows the predictive class matrix of KNN model.

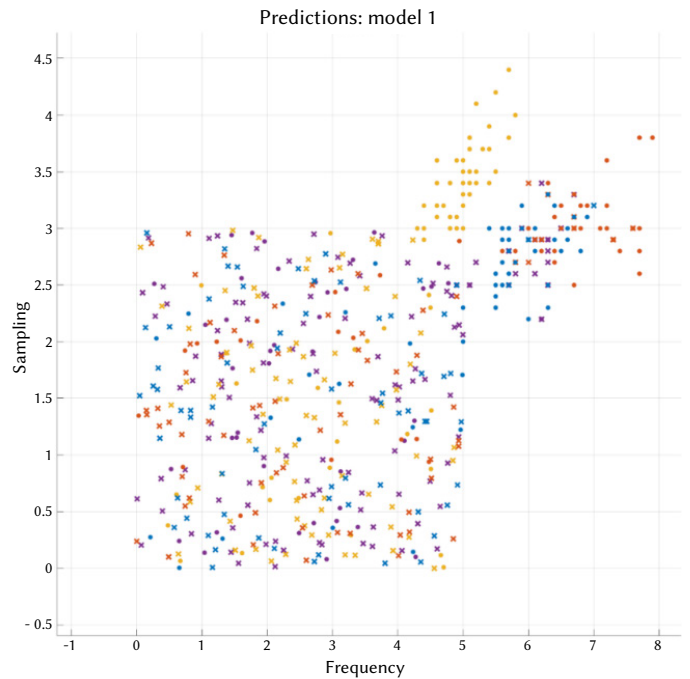


Fig. 5. The original training dataset for eye movement signals with features and manual labels.

		DGCNN					
True Class		'calm'	'happy'	'nervous'	'sad'	TPR	FNR
		Predicted Class	87.2%	3.2%	3.4%	6.0%	87.2%
'calm'		87.2%	3.2%	3.4%	6.0%	87.2%	12.8%
'happy'		2.5%	87.8%	3.7%	6.0%	87.8%	12.2%
'nervous'		5.3%	2.7%	88.8%	3.2%	88.8%	11.2%
'sad'		3.6%	5.2%	3.1%	88.1%	88.1%	11.9%

Fig. 6. Accuracy predictive matrix and AOC for DGCNN classification model.

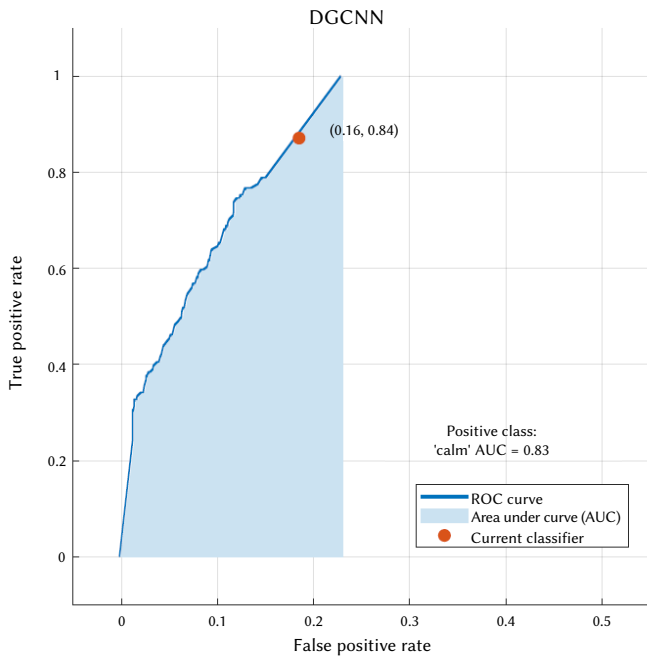


Fig. 7. ROC curvature for the TPR and FNR (affective =calm).

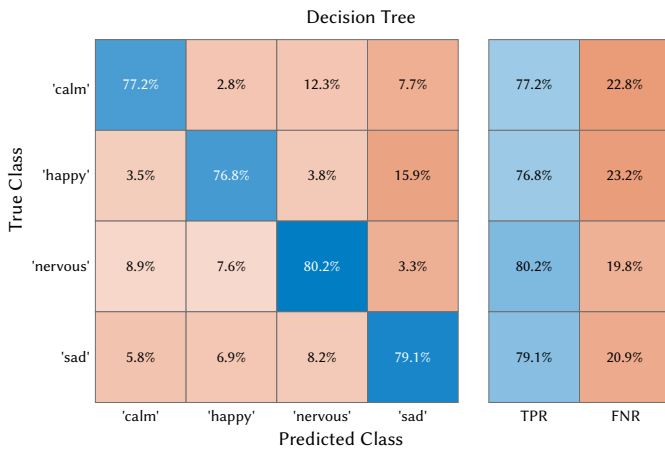


Fig. 8. Predictive class matrix of decision tree model for eye movement and tracking signals.

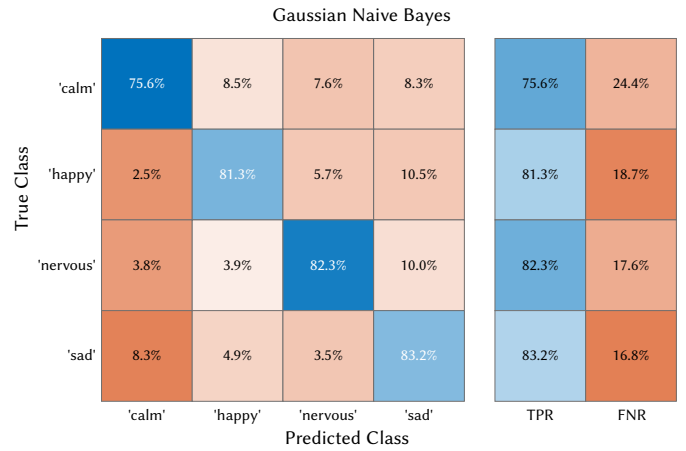


Fig. 9. Predictive class matrix of Gaussian naïve Bayes model for eye movement and tracking signals.

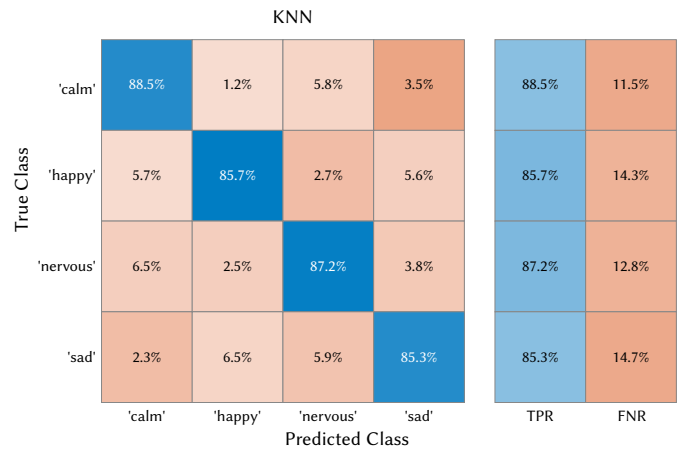


Fig. 10. predictive class matrix of KNN model for eye movement and tracking signals.

The matrix known that the proposed DGCNN performs a much higher predicted class in TPR and lower in FNR. Furthermore, more indices are developed for comparing the proposed model for classifying affective signals in which, A is accuracy; S1 is sensitivity; S2 is specific; PP is positive predictive; NP is negative predictive; PL is positive likely; NL is negative likely; F1 is f1-score which defined based on TPR and NPR. The comparing results are shown in Table I.

TABLE I. COMPARATIVE ANALYSIS ON THE DIFFERENT CLASSIFIERS AND THE PROPOSED CLASSIFICATION MODEL FOR ACOUSTIC SIGNAL ANALYSIS WITH STFT PREPROCESSED SIGNALS

Classifier/Evaluation Metrics	A	S1	S2	PP	NP	PL	NL	F1
ANN [43]	76.40%	81.23%	77.43%	78.12%	77.12%	7.21	0.13	0.91
MCSVM [44]	78.76%	82.12%	76.43%	76.32%	75.32%	7.22	0.12	0.82
LSTM [45]	81.42%	82.33%	76.78%	79.34%	77.32%	7.21	0.12	0.87
R-CNN [46]	82.43%	84.65%	78.76%	81.32%	78.12%	6.31	0.14	0.88
Fast R-CNN [47]	85.54%	84.43%	82.45%	82.57%	81.11%	6.54	0.12	0.87
VGG-16 [48]	85.43%	83.87%	84.67%	79.89%	81.81%	6.66	0.14	0.90
GoogLeNet [49]	87.11%	84.65%	83.55%	84.90%	81.78%	7.54	0.14	0.92
AlexNet [50]	85.76%	82.48%	85.12%	83.43%	82.32%	7.35	0.14	0.93
DGCNN (*)	88.10%	85.98%	82.69%	84.09%	84.87%	7.55	0.12	0.94



And the proposed DGCNN has still in high effectiveness in A, S1, NP, PL, and F1. Table I shows the comparing results for those indices, in which the DGCNN performs high effectiveness in most indices.

#### IV. CONCLUSION REMARKS

This article mainly studies the acquisition and preprocessing of eye movement signals, endpoint detection, effective eye movement signal extraction, feature extraction, and compares classification algorithms such as SVM, Bayes, and GoogleNet to verify the effectiveness of this method; eye movement analysis method is an effective method for studying human cognitive processing. Its advantages of harmless, ecological and high-efficiency are difficult to replace by general research technology; eye-tracking technology has become the foundation of psychology, neuromarketing, neurocognition, user experience technical methods for visual behavior and human behavior in many fields such as research and market research, recently, deep CNN was also used in sentiment analysis and emotion detection in conversations [51]. The research limitation is that the eye-tracking signals need to be converted and lose some information in feature extraction. Furthermore, future research requires multimodal physiological signals; on the other hand, more human emotion classifications also need to be further evaluated.

Eye movement signals can better characterize the emotional state and perform efficient classification; in this paper, the eye-tracking signals can be transformed into the standard input of the convolutional neural network by utilizing the eye movement signals' transformation and preprocessing. The effectiveness of the method is verified by comparison analysis. In future works, affective computing will be widely applied to medical rehabilitation, and assisting autistic people for their emotional changes, applying affective computing in education to realize the collection and analysis of learning status and guiding the selection of content progress. The computer can perceive the user's preference for music and based on the understanding and judgment of the emotional response, provide users with more interesting music playback.

#### REFERENCES

- [1] Rahal, R.-M. and S. Fiedler, "Understanding cognitive and affective mechanisms in social psychology through eye-tracking". *Journal of Experimental Social Psychology*, vol. 85, pp. 103842, 2019, doi: 10.1016/j.jesp.2019.103842
- [2] Kale, G. V., & Patil, V. H. "A study of vision based human motion recognition and analysis". *International Journal of Ambient Computing and Intelligence*, vol. 7, no.2, pp.75-92, 2016, doi: 10.4018/IJACI.2016070104
- [3] Gargava, P., & Asawa, K. Brain, "Computer Interface for Micro-controller Driven Robot Based on Emotiv Sensors", *International Journal of Interactive Multimedia & Artificial Intelligence*, vol. 4, no.5, pp.39-43, 2017, doi: 10.9781/ijimai.2017.457
- [4] Poria, S., Cambria, E., Bajpai, R., & Hussain, A., "A review of affective computing: From unimodal analysis to multimodal fusion", *Information Fusion*, vol. 37, pp. 98-125, 2017, doi: 10.1016/j.inffus.2017.02.003
- [5] Lee, W., & Norman, M. D. "Affective Computing as Complex Systems Science", *Procedia Computer Science*, vol. 95, pp. 18-23, 2016, doi: 10.1016/j.procs.2016.09.288
- [6] Belkacem, A.N., et al., "Erratum to Online classification algorithm for eye-movement-based communication systems using two temporal EEG sensors", *Control, Biomedical Signal Processing and Control*, vol. 19, pp. 137, 2015, doi: 10.1016/j.bspc.2015.01.006
- [7] Dai, W., Han, D., Dai, Y., et al., "Emotion recognition and affective computing on vocal social media", *Information & Management*, vol. 52, no.7, pp.777-788, 2015, doi: 10.1016/j.im.2015.02.003
- [8] Choi, M., Seo, M., Lee, J. S., et al., "Fuzzy support vector machine-based personalizing method to address the inter-subject variance problem of physiological signals in a driver monitoring system", *Artificial Intelligence in Medicine*, vol. 105, pp.101843, 2020, doi: 10.1016/j.artmed.2020.101843
- [9] R.C.A. Bendall, S. Lambert, A. Galpin, et al., "A cognitive style dataset including functional near-infrared spectroscopy, eye-tracking, psychometric and behavioral measures", *Data in Brief*, vol. 26, 104544, 2019, doi: 10.1016/j.dib.2019.104544
- [10] John Brand, Travis D. Masterson, et al., "Measuring attentional bias to food cues in young children using a visual search task: An eye-tracking study", *Appetite*, vol. 148, 104610, 2020, doi: 10.1016/j.appet.2020.104610
- [11] Wen-ying Sylvia Chou, Neha Trivedi, et al., "How do social media users process cancer prevention messages on Facebook? An eye-tracking study", *Patient Education and Counseling*, vol. 103, no.6, pp.1161-1167, 2020, doi: 10.1016/j.pec.2020.01.013
- [12] Gisele C. Gotardi, Sérgio T. Rodrigues, Fabio A. Barbieri, et al., "Wearing a head-mounted eye tracker may reduce body sway", *Neuroscience Letters*, vol. 722, 134799, 2020, doi: 10.1016/j.neulet.2020.134799
- [13] Hessels, R.S. and I.T.C. Hooge, "Eye tracking in developmental cognitive neuroscience - The good, the bad and the ugly", *Developmental Cognitive Neuroscience*, vol. 40, 100710, 2019, doi: 10.1016/j.dcn.2019.100710
- [14] Juhola, M., H. Aalto, and T. Hirvonen, "Using results of eye movement signal analysis in the neural network recognition of otoneurological patients", *Computer Methods and Programs in Biomedicine*, vol. 86, no.3, pp. 216-226, 2017, doi: 10.1016/j.cmpb.2007.02.008
- [15] Juhola, M., T. Tossavainen, and H. Aalto, "Influence of lossy compression on eye movement signals", *Computers in Biology and Medicine*, vol. 34, no.3, pp.221-239, 2004, doi: 10.1016/S0010-4825(03)00059-3
- [16] Enkelejda Kasneci, Thomas Kübler, Klaus Broelemann, et al., "Aggregating physiological and eye tracking signals to predict perception in the absence of ground truth", *Computers in Human Behavior*, vol. 68, pp. 450-455, 2017, doi: 10.1016/j.chb.2016.11.067
- [17] Oliver Faust, Yuki Hagiwara, Tan Jen Hong, et al., "Deep learning for healthcare applications based on physiological signals: A review". *Computer Methods and Programs in Biomedicine*, vol. 161, pp.1-13, 2018, doi: 10.1016/j.cmpb.2018.04.005
- [18] Jue Li, Heng Li, Waleed Umer, et al., "Identification and classification of construction equipment operators' mental fatigue using wearable eye-tracking technology", *Automation in Construction*, vol.109, 103000, 2020, doi: 10.1016/j.autcon.2019.103000
- [19] Xue Wang, Lin Lin, Meiqi Han, et al., "Impacts of cues on learning: Using eye-tracking technologies to examine the functions and designs of added cues in short instructional videos", *Computers in Human Behavior*, vol. 107, 106279, 2020, doi: 10.1016/j.chb.2020.106279
- [20] Dexiang Zhang, Jukka Hyönä, Lei Cui, et al., "Effects of task instructions and topic signaling on text processing among adult readers with different reading styles: An eye-tracking study", *Learning and Instruction*, vol. 64, 101246, 2019, doi: 10.1016/j.learninstruc.2019.101246
- [21] Król, M.E. and M. Król, "A novel machine learning analysis of eye-tracking data reveals suboptimal visual information extraction from facial stimuli in individuals with autism", *Neuropsychologia*, vol.129, pp.397-406, 2019, doi: 10.1016/j.neuropsychologia.2019.04.022
- [22] Angelina Vernetti, Atsushi Senju, Tony Charman, "Simulating interaction: Using gaze-contingent eye-tracking to measure the reward value of social signals in toddlers with and without autism", *Developmental Cognitive Neuroscience*, vol. 29, pp.21-29, 2018, doi: 10.1016/j.dcn.2017.08.004
- [23] Vettori, S., et al., "Combined frequency-tagging EEG and eye tracking reveal reduced social bias in boys with autism spectrum disorder", *Cortex*, vol.125, pp.135-148, 2020, doi: 10.1016/j.cortex.2019.12.013
- [24] Ding, X., & Lv, Z., "Design and development of an EOG-based simplified Chinese eye-writing system", *Biomedical Signal Processing and Control*, vol. 57, 101767, 2020, doi: 10.1016/j.bspc.2019.101767
- [25] Dua, M., Gupta, R., Khari, M., et al., "Biometric iris recognition using radial basis function neural network", *Soft Computing*, vol. 23, no. 22, pp.11801-11815, 2019, doi: 10.1007/s00500-018-03731-4
- [26] Dutta, A., Mondal, A., Dey, N., et al., "Vision Tracking: A Survey of the State-of-the-Art", *SN Computer Science*, vol.1, no.1, pp. 57, 2020, doi:10.1007/s42979-019-0059-z
- [27] Khari, M., Garg, A. K., Crespo, R. G., et al., "Gesture Recognition of RGB and RGB-D Static Images Using Convolutional Neural Networks", *International Journal of Interactive Multimedia & Artificial Intelligence*, vol. 5, no. 7, pp.22-27, 2019, doi: 10.9781/ijimai.2019.09.002
- [28] Ahuja, R., Jain, D., Sachdeva, D., et al., "Convolutional Neural Network

- Based American Sign Language Static Hand Gesture Recognition”, *International Journal of Ambient Computing and Intelligence*, vol. 10, no.3, pp. 60-73, 2019, doi: 10.4018/IJACI.2019070104
- [29] Deng, M., Meng, T., Cao, J., et al., “Heart sound classification based on improved MFCC features and convolutional recurrent neural networks”, *Neural Networks*, vol.130, pp. 22-32, 2020, doi: 10.1016/j.neunet.2020.06.015
- [30] Raj, R., Rajiv, P., Kumar, P., et al., “Feature based video stabilization based on boosted HAAR Cascade and representative point matching algorithm”, *Image and Vision Computing*, vol. 101, 103957, 2020, doi: 10.1016/j.imavis.2020.103957
- [31] Dey, N., Ashour, A. S., & Hassanien, A. E., “Feature detectors and descriptors generations with numerous images and video applications: a recap”, *Feature detectors and motion detection in video processing*, 2017, pp. 36-65, IGI Global.
- [32] Zhou, X., et al., “Eye tracking data guided feature selection for image classification”, *Pattern Recognition*, vol. 63, pp. 56-70, 2017, doi: 10.1016/j.patcog.2016.09.007
- [33] Ali, M. N. Y., Sarowar, M. G., Rahman, M. L., et al., “Adam deep learning with SOM for human sentiment classification”, *International Journal of Ambient Computing and Intelligence*, vol. 10, no. 3, pp. 92-116, 2019, doi: 10.4018/IJACI.2019070106
- [34] Shah, S., Kumar, A., Kumar, R., & Dey, N., “A robust framework for optimum feature extraction and recognition of P300 from raw EEG”, *U-Healthcare Monitoring Systems*, 2019, pp. 15-35, Academic Press.
- [35] Dey, A., Bhattacharya, D. K., Tibarewala, D. N., et al., “Chinese-chi and Kundalini yoga Meditations Effects on the Autonomic Nervous System: Comparative Study”, *International Journal of Interactive Multimedia & Artificial Intelligence*, vol. 3, no. 7, pp. 87-95, 2016, doi:10.9781/ijimai.2016.3713
- [36] Wu, H., & Zhao, X., “A Small and Portable Foot Motion Recognition Device Used in VR Environment”, *International Journal of Ambient Computing and Intelligence*, vol. 10, no.3, pp. 1-16, 2019, doi: 10.4018/ijaci.2019070101
- [37] Gopal, B., & Manohar, S., “VLSI architecture for the Winograd Fourier transform algorithm”, *Microprocessing and Microprogramming*, vol. 40, no. 9, pp. 605-616, 1994, doi: 10.1016/0165-6074(94)90089-2
- [38] Prabhakar, D. V. N., Sreenivasa Kumar, M., & Gopala Krishna, A., “A Novel Hybrid Transform approach with integration of Fast Fourier, Discrete Wavelet and Discrete Shearlet Transforms for prediction of surface roughness on machined surfaces”, *Measurement*, vol. 164, 108011, 2020, doi: 10.1016/j.measurement.2020.108011
- [39] George, A. E. W., So, S., Ghosh, R., & Paliwal, K. K., “Robustness metric-based tuning of the augmented Kalman filter for the enhancement of speech corrupted with coloured noise”, *Speech Communication*, vol. 105, 62-76, 2018, doi: 10.1016/j.specom.2018.10.002
- [40] Olbrys, J., & Mursztyn, M., “Estimation of intraday stock market resiliency: Short-Time Fourier Transform approach”, *Physica A: Statistical Mechanics and its Applications*, vol. 535, 122413, 2019, doi:
- [41] Mateo, C., & Talavera, J. A., “Short-Time Fourier Transform with the Window Size Fixed in the Frequency Domain (STFT-FD): Implementation”, *SoftwareX*, vol. 8, pp.5-8, 2018, doi: 10.1016/j.physa.2019.122413
- [42] Y. LeCun, L. Bottou, Y. Bengio et al., “Gradient-Based Learning Applied to Document Recognition”, *Proceedings of the IEEE*, vol. 86, no.11, pp.2278-2324, 1998, doi: 10.1109/5.726791
- [43] Hande Erkamaz, Mahmut Ozer, İlhami Muharrem Orak, “Detection of directional eye movements based on the electrooculogram signals through an artificial neural network”, *Chaos, Solitons & Fractals*, vol. 77, pp. 225-229, 2015, doi: 10.1016/j.chaos.2015.05.033
- [44] Bo Liu, Yanshan Xiao, Longbing Cao, “SVM-based multi-state-mapping approach for multi-class classification”, *Knowledge-Based Systems*, vol. 129, pp.79-96, 2017, doi: 10.1016/j.knosys.2017.05.011
- [45] Ying Wang, Qun Wu, Nilanjan Dey, et al., “Deep back propagation-long short-term memory network based upper-limb sEMG signal classification for automated rehabilitation”, *Biocybernetics and Biomedical Engineering*, vol. 40, no.3, pp. 987-1001, 2020, doi: 10.1016/j.bbe.2020.05.003
- [46] Yu Wang, Yating Chen, Ningning Yang, et al., “Classification of mice hepatic granuloma microscopic images based on a deep convolutional neural network”, *Applied Soft Computing*, vol. 74, pp. 40-50, 2019, doi: 10.1016/j.asoc.2018.10.006
- [47] Dan Wang, Zairan Li, Nilanjan Dey, “Optical pressure sensors based

plantar image segmenting using an improved fully convolutional network”, *Optik*, vol. 179, 99-114, 2019, doi: 10.1016/j.ijleo.2018.10.155

- [48] Zhenzhen Song, Longsheng Fu, Jingzhu Wu, et al., “Kiwi fruit detection in field images using Faster R-CNN with VGG16”, *IFAC-PapersOnLine*, vol. 52, no. 30, pp. 76-81, 2019, doi: 10.1016/j.ifacol.2019.12.500
- [49] Pengjie Tang, Hanli Wang, Sam Kwong, “G-MS2F: GoogLeNet based multi-stage feature fusion of deep CNN for scene recognition”, *Neurocomputing*, vol. 225, pp. 188-197, 2017, doi: 10.1016/j.neucom.2016.11.023
- [50] Ashkan Shakarami, Hadis Tarrah, AliMahdavi-Hormat, “A CAD system for diagnosing Alzheimer’s disease using 2D slices and an improved AlexNet-SVM method”, *Optik*, vol. 212, 164237, 2020, doi: 10.1016/j.ijleo.2020.164237
- [51] M. G. Huddar, S. S. Sannakki, V. S. Rajpurohit, “Attention-based Multimodal Sentiment Analysis and Emotion Detection in Conversation using RNN”, *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 6, 2020, doi:10.9781/ijimai.2020.07.004



Yuanfeng Li

He received his B.Eng. degree from Zhejiang A&F University in 2012, Master of Design Science in 2015 from Zhejiang Sci-Tech University. In 2016, he was appointed as a Lecturer at Jiyang College of Zhejiang A&F University. His research interests include human factors engineering and human-computer interaction focusing on user experience designs.



Jiangang Deng

He received Bachelor of Arts from Zhejiang A&F University in 2003, Master of Agriculture from Beijing Forestry University in 2009. In 2008, he has appointed as a Lecturer at Zhejiang A&F University. His topic research is product innovation design, human interactive design, intelligent design, green design and computer aided conceptual design.



Qun Wu

Dr. 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 a M.E. degree in Mechanical Engineering from Shaanxi University of Science and Technology, China, in 2004. His research interests include machine learning, human factor and product innovation design.



Ying Wang

She received the B.Eng. degree in Digital Art and Design from Zhejiang University, China in 2014. She is now employed as an Associate Professor in Department of Industrial Design at College of Art and Design, Zhejiang Sci-tech University, Hangzhou, China. Her research topic is universal design, HCI for elderly, biosensor based technologies for interactive design, rehabilitation engineering design, and computer aided intelligent design.