International Journal of Gaming and Computer-Mediated Simulations
Volume 14 • Issue 2

Feature Extraction Method of Piano Performance Technique Based on Recurrent Neural Network

Zhi Qian, Baoji University of Arts and Sciences, China*

ABSTRACT

In order to solve the problem of low efficiency in traditional feature extraction methods of piano performance techniques, a feature extraction method of piano performance techniques based on recurrent neural network is proposed. Analyze the types of piano playing techniques, and establish the hand model. On this basis, the hand action signals of piano performance are collected from the two aspects of finger key strength and hand action video image. Finally, the feature extraction of piano performance techniques is realized from the time domain and frequency domain. Through the comparison with the traditional extraction method, it is concluded that the extraction efficiency of the optimized design of piano performance technique feature extraction method has been significantly improved, and it has obvious application advantages in the identification of piano performance techniques.

KEYWORDS

feature extraction, recurrent neural network, performance techniques, piano performance

INTRODUCTION

Piano performance carries the emotional and essential connotation of piano works, which needs to be expressed through the performer's second creation, so as to arouse the audience's psychological resonance. The process of piano playing fully integrates skills and emotions, and expresses and deduces the inner beauty of music. The analysis from the form of piano playing technology can provide good technical guidance and aesthetic guidance for performers. Piano playing, as a form of artistic expression, has a high degree of freedom and openness. It is not only the performance of the player's own piano playing technology, but also the embodiment of the player's second creation of piano music works, and also the manifestation of the player's pursuit of art. Different from other ways of playing, piano playing has strict technical requirements. Specifically, playing requires not only the expression of the performer's basic psychological reflection on the music, but also the expression of the performer's aesthetic feelings on the music works (Raghuwanshi et al., 2019). Therefore, in order

DOI: 10.4018/IJGCMS.314589

*Corresponding Author

to obtain the perfect performance effect, we must deeply grasp the connotation of piano performance technology. Piano playing technique refers to the whole process of expressing the music symbols of music score with real songs. In this process, the piano player will convey the emotion, attitude and values to the audience. Therefore, the performer realizes his own second artistic creation through piano playing technology. Piano performance technology is the performer's deep understanding and accurate interpretation of artistic works. A piano concerto is a form of a symphony that is created for a pianist and is often supported through orchestras or another big group in the traditional music styles. Piano concertos are often virtuoso showcases that need a high level of technical proficiency on the device. Several sectors, particularly student learning as well as computerized behavior modeling, are interested in the quantitative assessment of piano playing. The intricate qualities of free concerts, in which the performer's professional capabilities are exhibited in harmony with his or her personal perception of the piece, make a systematic estimate of a piano performance extremely challenging. Piano playing technology is reflected on the surface through the coordinated movement of the player's fingers and other parts. Therefore, piano playing technology is not only in the scope of music performance, but also in the scope of sports science. Piano playing technology can be implemented perfectly by effectively controlling the change of timbre and music (Zhang et al., 2019).

In order to provide sufficient teaching resources for piano performance career, this paper uses computer technology to extract the characteristics of piano performance techniques, obtains the performance rules through the analysis of the extracted characteristics, and improves the performance level of piano learners. Feature extraction is a concept in computer vision and image processing. The technique of translating unprocessed signals into digital characteristics that can be handled whilst keeping the data into a single collected data is known as feature extraction. It produces positive performance by performing machine learning to basic information automatically. The average of a signal's window is an essential responsibility characteristic. It refers to the use of computer to extract image information, determine whether each image point belongs to an image feature. The result of feature extraction is to divide the points on the image into different subsets, which often belong to isolated points, continuous curves or continuous regions. Therefore, the feature extraction of piano performance techniques is based on the video image of piano performance to get the features reflecting the performance techniques (Hiranmai et al., 2019). Experienced musicians use more proximate movements, gravitation, inter-segment, as well as response pressures than beginner pianists to minimize neuromuscular stress but also attain the physiological performance of exhaustion tissue. Applicable in scenarios with a lot of dimensions. It saves memory by using a selection of training images termed network parameters in the classification model. For the determination algorithms, several optimization techniques can be given, as well as bespoke kernels. Based on the research status at home and abroad, the mature feature extraction methods include feature extraction based on correlation analysis, feature extraction based on support vector machine and feature extraction based on AdaBoost algorithm. AdaBoost is a machine learning method that may be utilized to increase the efficiency of any other machine learning technique. It works well with students who are struggling. On supervised learning, these are systems that reach efficiency slightly over random chance. Decision trees by one stage are the most suitable and hence more commonly used method with AdaBoost. AdaBoost's most significant benefits are Low generalization error, simple to build, supports a wide range of categories, and requires no adjustment of variables. Because this technique is sensitive to noise, special emphasis on information is necessary. However, in the actual extraction process, the traditional extraction method has the problems of slow extraction speed and incomplete extraction results. When it is applied to the recognition of piano playing techniques, the accuracy of the recognition results is low. The concept of recurrent neural network is introduced. Recurrent neural network is the general name of two kinds of artificial neural network. One is time recurrent neural network; the other is structure recurrent neural network. It is only effective in time series prediction since it has the ability to recall past information. This is referred to as Long Short-Term Memory (LSTM). Even CNN models are employed with RNN to broaden the functional visual neighborhood.

Over time, an RNN retains each bit of knowledge. It is only effective in function approximation since it has the ability to recall past inputs. The connection between neurons of time recurrent neural network constitutes a directed graph, while structural recurrent neural network uses similar neural network structure to recursively construct more complex depth network. It employs similar settings for every source since it produces the equivalent outcome by performing the identical operation on various entries or hidden nodes. Unlike other neural networks, this decreases the sophistication of the variables. The two training algorithms are different, but they belong to the same algorithm variant. In this paper, recurrent neural network is applied to the feature extraction of piano playing techniques. Through the learning and iteration of neural network, the image signal is classified and processed, which indirectly improves the integrity of feature extraction results.

DESIGN OF FEATURE EXTRACTION METHOD FOR PIANO PERFORMANCE TECHNIQUES

The Types of Piano Playing Techniques

Piano playing techniques include two hands crossing and long-distance jump, two hands fast alternating round playing technique, fast scale, fast repetition single finger technique, jump in, parallel interval, tremolo, arpeggio and so on. Big jump refers to two strides of more than 8 degrees, which requires piano players to accurately remember the position of the octave playing arpeggio (Suresha et al., 2020); (Mezzoudj et al., 2018). Take the music work K.120 of the British composer and viola player John Bull as an example. During the performance of the piano work, the left hand constantly strides to the highest range of the keyboard, and the corresponding right hand naturally strides to the lowest range of the keyboard, and sometimes both hands at the same time. The technique of quick alternation of hands is mainly to train the technique of quick alternation of hands, so as to make the single melody more cheerful and lively. It is required to make the timbre bright and the grain clear when the hands are fast alternation of hands. Jump in performance technique is to play fast and repeated jump by moving both arms back and forth, in order to replace the original only relying on finger playing (Ameur et al., 2020). In addition, the arpeggio technique in piano performance includes 24 major and minor arpeggios and 7 minus 7 arpeggios. This is a scale-based concept that focuses on the ability to expand and contract between fingers. In piano music, the arpeggio part accounts for a large part, and all kinds of arpeggios will appear, but they are all 24 arpeggio deformation modes with 7 major and 7 minor chords.

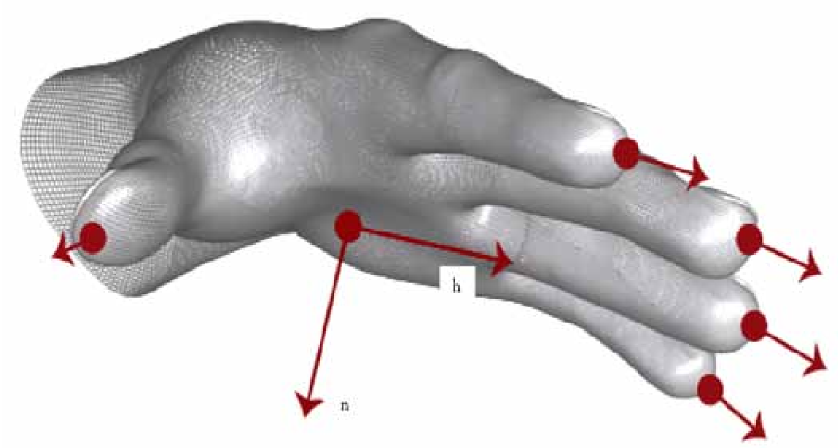
Building Hand Model

The feature extraction of piano playing techniques is based on the hand model to identify the finger direction vector, and the model is constantly verified by the data collected by two ultra-wide angle infrared sensors in real time (Ai et al., 2019); (Li. et al., 2020). By collecting the player’s hand joint angle parameters, phalange length, palm scale, arm length and other linear parameters, the hand model is established, and the results are shown in Figure 1.

COLLECTING THE HAND MOVEMENT SIGNAL OF PIANO PLAYER

The motion signals are collected from two aspects of finger key strength and hand motion video image. The finger key strength signal mainly uses pressure sensor equipment, which combines silicon piezoresistive sensor, microcomputer and digital signal processing technology, be able to feel the specified measured signal and convert it into available output signal according to certain conversion rules (Shanableh et al., 2020). Piezoresistive sensors can detect changes in electromagnetic impedance in highly conductive substances like silicone to calculate pressure. Predictability, concise overview circuitry (for translating impedance into a matching voltage applied), easy production process, cheap

Figure 1. Schematic diagram of hand model



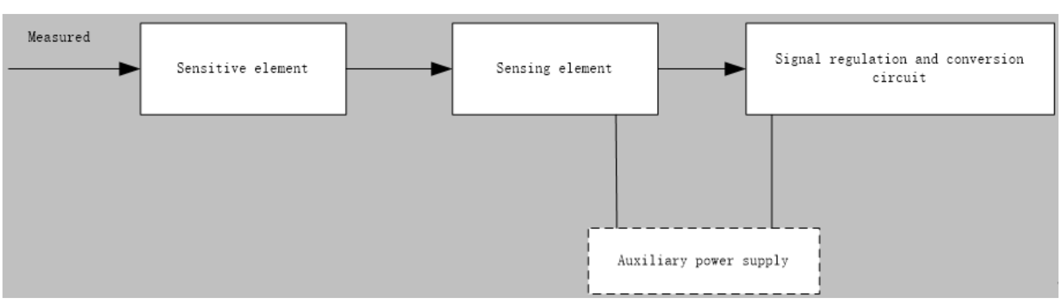
U cost, inexpensive measuring technique, a diverse variety of pressure transducers, as well as excellent precision are all benefits of the piezoresistive sensing element. The structure of the pressure sensor is shown in Figure 2.

The measured values are carried out by a sensitive element to the sensing element. Then, signal regulation is used and converted the signals by using an appropriate circuit. In between auxiliary power supply of the pressure sensor is observed. The prepared sensor equipment is installed on each key of the piano. By reading the pressure data of the key at different positions, the collection results of the signal data of finger key strength in the process of piano playing are obtained. In addition, the main application of hand motion video image is camera equipment. The camera is installed on the piano music stand to ensure that all hand motion information can be collected.

ORIGINAL SIGNAL PREPROCESSING

Before feature extraction, the initial signal is preprocessed. Through the use of filtering technology, we can extract the required content from the relatively complex background signal, and suppress the unnecessary content. For engineering practice, the filter is a kind of frequency selective device. When the signal passes through the filter, it can make the required electrical signal of a certain frequency have a small attenuation, so that this part of the required signal can smoothly pass through the filter, and at the same time make other unnecessary electrical signals have a large attenuation, to a certain

Figure 2. Composition and structure of pressure sensor



extent, prevent these useless signals from passing through the filter (Hua et al., 2019). The specific filtering process can be expressed as follows:

$$H\left(z\right)=\sum_{n=0}^{N-1}h\left(n\right)z^{-n} \tag{1}$$

In the formula, $h\left(n\right)$ is the initial acquisition signal and z^{-n} is the signal frequency. When it comes to estimating the dynamic response of fluctuating impulses, the Continuous Wavelet Transform (CWT) is quite effective (for instance, detection of damping in dynamic networks). CWT is also extremely robust to additive noise. Then, continuous wavelet transform is used to de noise the initial signal:

$$H\left(t\right)=\left|a\right|^{-\frac{1}{2}}H\left(\frac{t-b}{a}\right) \tag{2}$$

In the formula, a and b set the upper and lower limits of threshold for wavelet transform respectively. Inside a total population, Gaussian Mixture modeling is utilized to describe Normally Distributed population groups. Heterogeneous approaches have the benefit of not requiring which subgroup an information piece corresponds to. It enables the algorithm to extract knowledge from the different populations. On this basis, the piano player's hand area is detected, and the adaptive Gaussian mixture model background construction method is adopted. The gray value distribution of each pixel is represented by a mixture of multiple Gaussian distributions (Triwiyanto et al., 2020); (Ding et al., 2015). If the pixel observation value of a pixel is X_r, the probability of the current pixel value can be expressed as:

$$p\left(X_t\right)=\sum_{i=1}^{k}\omega\left(i,t\right)\eta\left(X_t,u\left(i,t\right),\sum i,t\right) \tag{3}$$

In the formula, k is the number of Gaussian distributions used to represent pixel values, $\omega\left(i,t\right)$ is the weight of the i th Gaussian distribution at time t, $u\left(i,t\right)$ is the mean value of the i th Gaussian distribution, $\sum i,t$ is the covariance matrix of the i th Gaussian distribution, and the expression of the i th Gaussian distribution function $\eta\left(X_t,u\left(i,t\right),\sum i,t\right)$ is as follows:

$$\eta\left(X_t,u\left(i,t\right),\sum i,t\right)=\frac{1}{\left(2\pi\right)^{\frac{n}{2}}\sqrt{\left|\sum i,t\right|}}e^{-\frac{1}{2}\phi} \tag{4}$$

In the formula, $\sum i,t$ is the covariance matrix and n is the number of free variables. The pixel value of each point in the first image is used as the mean value of the corresponding Gaussian mixture distribution, and each Gaussian model is given a larger variance and a smaller weight (Johnson et al., 2020). When the scene changes, the Gaussian mixture model of each pixel needs to be continuously learned and updated. A Gaussian mixture model is a classification algorithm in which all pieces of information are created by combining a set amount of Gaussian distributions with uncertain variables. One or several Gaussian distributions of each point are selected as the background model, and the others are foreground models. If the current value matches the background model, the point is the background, otherwise it is the foreground (Lostanlen et al., 2021). Through the segmentation of foreground and background, the hand region is detected and extracted, and it is continuously tracked.

SETTING THE FEATURE EXTRACTION ITEM OF PIANO PERFORMANCE TECHNIQUE

Setting the key sequence, key strength, rhythm and pedal, as the feature extraction items of piano performance techniques. The key sequence of piano performance is to record the key situation of piano research according to the time sequence of video. The key strength can be directly obtained by reading the data from the pressure sensor. According to the key strength, the piano playing techniques can be divided into three types: weight, delicacy and singing (Gautam et al., 2020). For example, the use of song playing method is the performance means of music singing. It requires that when performing the singing melody, we should stick the key with our fingers and ensure the clarity of the music. At the same time, we should pay attention to the transfer of finger power and the stability of fingertips. The weight playing method maintains the minimum distance between the fingers and the keyboard, so that the fingers and wrists can be relaxed. For the delicate playing method, the pursuit is light and weightless playing, and the feeling of upper and lower arms should be light and relaxed (Miljkovi et al., 2021), (Yan et al., 2019). You can press the key to the end, not to withdraw the key upward with your fingers, but to maintain it on the surface of the key, so that the power on the key can be shared and the unevenness of timbre can be avoided. The rhythm of playing is the time interval between two notes. According to the time interval, the piano playing rhythm can be divided into three types: fast rhythm, medium rhythm and slow rhythm. In addition, the use of the pedal is to record the frequency, position and way of using the pedal in the process of playing.

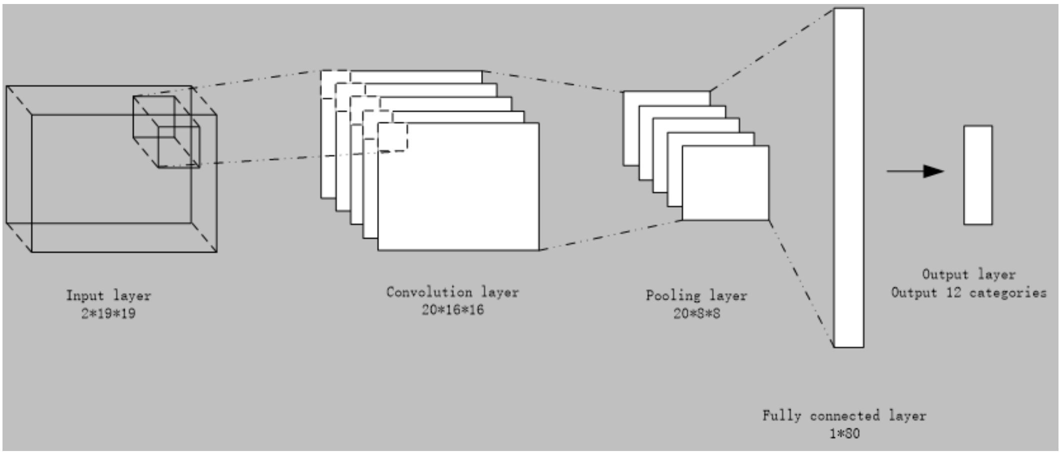
CLASSIFICATION OF HAND ACTION SIGNAL TYPES USING RECURRENT NEURAL NETWORK

A recurrent neural network is established, and the processed piano playing signal is taken as the input to realize the classification of hand action signal (Krishnamurthi et al., 2019); Lokesh, et al., 2018). Nevertheless, due to its system storage, a recurrent neural network may recall such characteristics. It generates output, replicates it, as well as then feeds it back into the connection. Simply, recurrent neural networks combine historical information with information from the contemporary. The established recurrent neural network adopts a five-layer network structure, including an input layer, a convolution layer, a maximum pooling layer, a full connection layer and an output layer. The network structure is shown in Figure 3.

The feature maps' dimensionality is reduced by using pooling layers. As a result, the set of indicators to acquire as well as the quantity of processing in the system is both reduced. The result from the completed Consolidation or Feedforward neural Layer, which is flattening and also supplied into the deep network, is the intake to the perceptron. Computational complexity has been used in computer vision for a considerable period to distort as well as enhance pictures, but they may also be employed to do other tasks. CNNs impose a specific line network among synapses of neighboring levels (i.e., increase borders and embossed).

A recurrent neural network (RNN) is a form of artificial neural network which is designed to deal with response variables or periods of information. RNNs feature a conception of memory,' which allows systems to retain the conditions or knowledge of prior outputs in order to construct the sequence's upcoming outcome. In the recurrent neural network shown in Figure 3, the key convolution operation is completed in the convolution layer. The number of operations convoluted with the input characteristic graph is the convolution kernel, and its parameters include width, length and depth. These parameters are trained by the network model, and each convolution region of the input feature graph shares the same set of weights and bias parameters in convolution operation (Xiao et al., 2020).

Figure 3. Structure diagram of recurrent neural network



When the convolution operation is completed, the bias parameter and nonlinear function are also needed to act on the convolution result, and then the output characteristic graph is obtained. The convolution operation in convolution neural network is different from the above convolution operation. The calculation process of convolution layer is shown in Formula (5).

$$y_j^{(l)} = f\left(\sum_{i \in K_j} y_j^{(l-1)} \times W_{ij}^{(l)} + B_j^{(l)}\right) y_j^{(l)} = f\left(\sum\nolimits_{i \in K_j} y_j^{(l-1)} \times W_{ij}^{(l)} + B_j^{(l)}\right) \quad (5)$$

In the formula, the superscript $\left(l\right)$ is the first network layer; The subscript *i* represents the *i*-th input characteristic graph and the subscript *j* represents the *j*-th output characteristic graph; $y_j^{(l)}$ represents the *j*-th characteristic graph obtained by convolution operation of the first layer, and *f* represents the activation function of the nonlinear layer. $y_j^{(l-1)}$ is the input characteristic graph of the layer, and it is also the $\left(l-1\right)$ output characteristic graph. Since the output characteristic graph is associated with each input characteristic graph, $W_{ij}^{(l)}$ is the convolution kernel between the *i*-th input characteristic graph and the *j*-th output characteristic graph in the 1 layer, and $B_j^{(l)}$ is the offset parameter on the *j*-th characteristic graph (Wang et al., 2019). Accumulation, also known as pool analysis or combined screening, is the process of mixing the identical kind of material from multiple persons and running a single NAAT lab experiment on the aggregated group of samples to identify SARS-CoV-2, the viral disease COVID-19. Pooling operation is to divide the feature graph into several sub regions, aggregate statistics and select a point as the output. There are three pooling methods: maximum pooling, average pooling and random pooling, which depend on the maximum value, average value and probability value for sampling respectively (Rastgoo et al., 2021). In practical application, in order to improve the nonlinear ability of network model, activation function is often introduced to form nonlinear layer, the activation function expression is:

$$Sigmoid\left(x\right) = \frac{1}{1+e^{-x}} \quad (6)$$

After a series of convolution layers and pooling layers, the input image is gradually mapped into feature vectors. Finally, the fully connected layer is used for classification. The full connection layer is equivalent to designing the same convolution kernel as the last layer of feature image, and reducing the dimension of two-dimensional feature image to get one-dimensional feature vector and fusion (Rastgoo et al., 2020).

REALIZING THE FEATURE EXTRACTION OF PIANO PERFORMANCE TECHNIQUES

Integrate all set piano playing techniques to extract feature items, and realize feature extraction from time domain and frequency domain respectively. The time-domain eigenvalues mainly include mean value, amplitude histogram, zero crossing times, root mean square value, variance, fourth-order origin matrix, autocorrelation function and so on, the feature extraction expression of root mean square value is as follows:

$$RMS_{(n)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} sEMG_i^2} \tag{7}$$

In the formula, the parameter $sEMG_i$ represents the amplitude of the signal at the *i*-th sampling, and *n* represents the number of samples. The frequency domain characteristics of signal are mainly analyzed by power spectrum method. The frequency eigenvalues include average power frequency and median frequency, the specific expression is as follows:

$$\begin{cases} MPF = \dfrac{\int_0^{\infty} f \cdot PSD(f)\, df}{\int_0^{\infty} PSD(f)\, df} \\ MF = \dfrac{1}{2}\int_0^{\infty} PSD(f)\, df \end{cases} \tag{8}$$

The intensity of fluctuations (energetic) as a function of wavelength is represented by the power spectral density function (PSD). In other words, it displays whether wavelengths have high fluctuations as well as whose harmonics have tiny differences. Throughout its channel capacity, a message made up of several comparable frequency components will have a continual density function (PSD), as well as, the overall transmit power can be calculated as P = PSD BW. In the formula, $PSD(f)$ is the power spectral density function of EMG (Zhang, Z, et al, 2020). Through the extraction of time domain features and frequency domain features, we can get the feature extraction results of finger angle, finger palm distance and finger height in piano playing techniques, the results are as follows:

$$\begin{cases} A_i = \angle\left(x_i^{\pi} - C, h\right) \\ D_i = \dfrac{F_i - C}{S} \\ E_i = \dfrac{\operatorname{sgn}\left(\left(F_i - x_i^{\pi}\right) \cdot n\right) F_i - x_i^{\pi}}{S} \\ S = F_{middle} - C \end{cases} \tag{9}$$

In the formula, S represents the distance between the middle finger and the palm, C is the coordinate value of the palm center in the three-dimensional space relative to the controller, h is the normal vector of the palm center pointing to the finger, n is the vector of the direction perpendicular to the palm plane away from the palm, F_i is the three-dimensional coordinate of the fingertip detected by the controller, and F_{middle} is the fingertip coordinate of the middle finger. Features A, D and E comprehensively consider the angle and distance between finger and palm, so as to distinguish different gestures. Feature a represents the angle between the direction of finger extension and the axis of palm plane, which mainly describes the direction of each finger in palm plane. Feature D represents the distance from finger to palm, which mainly describes the position of fingertip relative to palm. Feature E represents the elevation distance of fingertip relative to palm plane, which mainly describes the degree of finger bending (Tran, D.P, et al, 2019), (Wang, X, et al, 2018). In the same way, we can get the extraction results of all other set feature items.

COMPARATIVE EXPERIMENTAL ANALYSIS

In order to test the extraction function and application performance of the feature extraction method of piano performance techniques based on recurrent neural network, a comparative experiment is designed. Caffe models are machine learning algorithms that work from start to finish. The net is a collection of strata linked by a computing graph - specifically, an acyclic network (DAG). Caffe handles all of the accounting for any DAG of levels, ensuring that both forward and reverse operations are accurate. The training and feature extraction of recurrent neural network mentioned in the design method are based on deep learning Caffe framework, and implemented on a desktop computer equipped with Intel Core i7-6700 CPU @ 3.40 GHz, 16 GB memory and Nvidia Geforce GTX Titan x graphics processor, while feature fusion and classification recognition are implemented by using MATLAB R2015b. The fundamental advantage of feature-level fusion is that it identifies a small set of prominent characteristics that might increase classification performance by detecting associated evaluation metrics provided by distinct biometric methods. In the virtual environment created by anaconda, this environment has python, C + + or C program library. Python was chosen because of its simplicity and compatibility with well-developed libraries supported by other major deep learning frameworks. TensorFlow is a Google-developed integrated development environment for running computer vision, pattern recognition, as well as other analytical and process system applications. It is a program that works with large datasets which are graphed as processing elements. Stream processing statistics that explain how to represent the information across a network or a sequence of computation elements be constructed with TensorFlow. In this project, TensorFlow framework is used to train sEMG data and recognize gesture action. The programming language is python. The operating system is windows, 4 GB memory, 1.5 GHz processor. The applications of TensorFlow framework are self-driving automobiles, systems for speech recognition, sentiment analysis is a technique for determining how people feel about something, recognition and labelling of images and videos, and text summarization is a useful skill to have. The initial parameter setting of recurrent neural network is shown in Table 1.

Collect the piano playing video in the network platform as the experimental technique to extract research samples. The preparation of some samples is shown in Figure 4.

The input video is uniformly processed into 32 frames, and the frame image size is changed to 128 * 171 after it is input to the network. Then each frame in the video is randomly clipped to 112 * 112. Gradient descent is a well-known design and implementation for training predictive models and neural nets. These systems eventually learn with the use of the dataset, as well as the objective functions inside steepest descent functions as a gauge, assessing its correctness with every repetition of variable parameters. The random gradient descent method is used to train the network: Firstly, the data is scrambled to reduce the interference of the data arrangement information on the training.

Table 1. Initial parameter setting table of recurrent neural network

Parameter number	Parameter name	Parameter value
1	Initial learning rate	0. 001
2	Training batch size	32
3	Dropout	0.5
4	Optimizer	A dam
5	Loss function	Cross entropy loss function
6	Offset value	0

Figure 4. Video sample of experimental piano performance

During each training iteration, there are two video input networks. The initial learning rate of the network is set to 0.0001, and it decreases by 10% after every 5000 iterations. The momentum is 0.9 and the weight decay term is 0.0005. The network stops training after 120000 iterations. Each time the random gradient decreases, the corresponding activation function is normalized by mini batch, so that the mean value of each dimension of the output signal is 0 and the variance is 1, so as to prevent the "gradient dispersion", accelerate the convergence speed and improve the accuracy of the model. Therefore, in the recurrent neural network, a BN layer will be connected after each convolution layer to achieve the effect of batch standardization.

In the experiment, the speed of feature extraction, extraction integrity and application recognition rate are set as the test indexes. In order to form the experimental comparison, the traditional feature extraction method and the feature extraction method based on AdaBoost algorithm are set as the two comparison methods of the experiment. The same experimental object is processed in the same experimental environment to ensure the uniqueness of the experimental variables. After the operation of three feature extraction methods and the statistics and processing of relevant data, the test results of feature extraction function are obtained, as shown in Table 2.

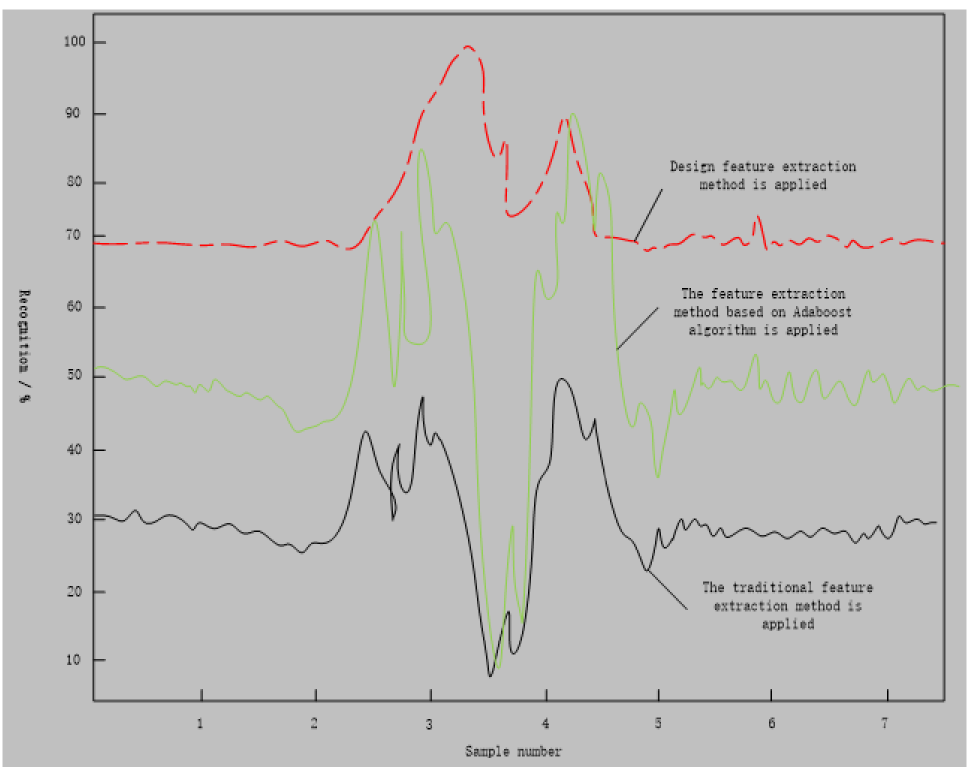
It can be seen from Table 2 that the average extraction time of the three feature extraction methods is 18.67 s, 15.87 s and 11.41 s respectively. In terms of feature extraction amount, the average extraction

Table 2. Test results of piano performance technique feature extraction function

Experimental group	Traditional feature extraction methods		Feature extraction method based on AdaBoost algorithm		Designed feature extraction method	
	Extraction time / s	Total extracted data / MB	Extraction time / s	Total extracted data / MB	Extraction time / s	Total extracted data / MB
1	18.35	0.49	15.69	0.66	11.23	0.73
2	18.61	0.51	15.37	0.65	12.12	0.81
3	18.27	0.46	15.28	0.62	10.89	0.77
4	18.44	0.48	16.04	0.58	11.53	0.84
5	19.03	0.52	16.22	0.59	11.68	0.75
6	19.12	0.54	16.13	0.61	12.04	0.78
7	19.05	0.46	15.97	0.65	10.92	0.84
8	18.47	0.47	16.22	0.57	10.84	0.82

amount of the three extraction methods is 0.49 MB, 0.62 MB and 0.79 MB respectively. It can be seen that the designed feature extraction method can extract more piano performance technique features in a shorter time, that is, the extraction efficiency is higher. The three feature extraction methods are applied to the recognition of piano playing techniques, and the corresponding recognition rate calculation and statistical results are shown in Figure 5.

Figure 5. Comparison curve of recognition rate of piano playing techniques



It can be seen from Figure 5 that the recognition rate of piano playing techniques corresponding to the feature extraction method of design is higher, that is, the application performance of the feature extraction method of design is better.

CONCLUSION

Piano playing technology has been extended from the original meaning of pure technology to the level of artistic expression, which is an important cognitive way of music culture. The change of piano performance technology is carried out under the background of the times, so it should be analyzed according to the background of the times. Under the support of recurrent neural network, using feature extraction method to achieve efficient analysis and extraction of piano performance techniques is of great significance for the development of piano performance education and art.

REFERENCES

Ai, X., Xu, Z., & Zhao, F. (2019). Feature extraction of micro-motional targets via time-range distribution. *IEEE Access: Practical Innovations, Open Solutions*, *7*(99), 1–1. doi:10.1109/ACCESS.2019.2937143

Ameur, S., Khalifa, A. B., & Bouhlel, M. S. (2020). Chronological pattern indexing: An efficient feature extraction method for hand gesture recognition with leap motion. *Journal of Visual Communication and Image Representation*, *70*, 102842. doi:10.1016/j.jvcir.2020.102842

Ding, X., & Zhang, R. (2015). Convergence of online gradient method for recurrent neural networks. *Journal of Interdisciplinary Mathematics*, *18*(1-2), 159–177. doi:10.1080/09720502.2015.1013721

Gautam, A., Panwar, M., Wankhede, A., Arjunan, S. P., Naik, G. R., Acharyya, A., & Kumar, D. K. (2020). Locomo-Net: A low -complex deep learning framework for sEMG-based hand movement recognition for prosthetic control. *IEEE Journal of Translational Engineering in Health and Medicine*, *8*, 2100812. doi:10.1109/JTEHM.2020.3023898 PMID:33014638

Hiranmai, M., Niranjana, K. B., & Nagaraj, H. K. (2019). Comparative study of various feature extraction techniques for pedestrian detection. *Procedia Computer Science*, *154*, 622–628. doi:10.1016/j.procs.2019.06.098

Hua, J., Li, G., Jiang, D., Zhao, H., & Qi, J. (2019). An optimized selection method of channel numbers and electrode layouts for hand motion recognition. *International Journal of HR; Humanoid Robotics*, *16*(04), 1941006. doi:10.1142/S0219843619410068

Johnson, D., Damian, D., & Tzanetakis, G. (2020). Detecting hand posture in piano playing using depth data. *Computer Music Journal*, *43*(1), 59–78. doi:10.1162/comj_a_00500

Krishnamurthi, R., Patan, R., & Gandomi, A. H. (2019). Assistive pointer device for limb impaired people: A novel Frontier Point Method for hand movement recognition. *Future Generation Computer Systems*, *98*, 650–659. doi:10.1016/j.future.2019.02.010

Li, X., & Geng, S. (2020). Research on sports retrieval recognition of action based on feature extraction and SVM classification algorithm. *Journal of Intelligent & Fuzzy Systems*, *39*(4), 5797–5808. doi:10.3233/JIFS-189056

Lokesh, S., Malarvizhi Kumar, P., Ramya Devi, M., Parthasarathy, P., & Gokulnath, C. (2019). An automatic tamil speech recognition system by using bidirectional recurrent neural network with self-organizing map. *Neural Computing & Applications*, *31*(5), 1521–1531. doi:10.1007/s00521-018-3466-5

Lostanlen, V., El-Hajj, C., Rossignol, M., Lafay, G., Andén, J., & Lagrange, M. (2021). Time–frequency scattering accurately models auditory similarities between instrumental playing techniques. *EURASIP Journal on Audio, Speech, and Music Processing*, *2021*(1), 1–21. doi:10.1186/s13636-020-00187-z PMID:33488686

Mezzoudj, F., & Benyettou, A. (2018). An empirical study of statistical language models: N-gram language models vs. neural network language models. *International Journal of Innovative Computing and Applications*, *9*(4), 189–202. doi:10.1504/IJICA.2018.095762

Miljković, N., & Isaković, M. S. (2021). Effect of the sEMG electrode (re) placement and feature set size on the hand movement recognition. *Biomedical Signal Processing and Control*, *64*, 102292. doi:10.1016/j.bspc.2020.102292

Raghuwanshi, G., & Tyagi, V. (2019). Impact of Feature Extraction Techniques on a CBIR System. In *International Conference on Advances in Computing and Data Sciences,* (pp. 338-348). Springer, Singapore. doi:10.1007/978-981-13-9939-8_30

Rastgoo, R., Kiani, K., & Escalera, S. (2020). Hand sign language recognition using multi-view hand skeleton. *Expert Systems with Applications*, *150*, 113336. doi:10.1016/j.eswa.2020.113336

Rastgoo, R., Kiani, K., & Escalera, S. (2021). Hand pose aware multimodal isolated sign language recognition. *Multimedia Tools and Applications*, *80*(1), 127–163. doi:10.1007/s11042-020-09700-0

Shanableh, T. (2021). Feature extraction and machine learning solutions for detecting motion vector data embedding in HEVC videos. *Multimedia Tools and Applications*, *80*(18), 27047–27066. doi:10.1007/s11042-020-09826-1

Suresha, M., Kuppa, S., & Raghukumar, D. S. (2020). A study on deep learning spatiotemporal models and feature extraction techniques for video understanding. *International Journal of Multimedia Information Retrieval*, *9*(2), 81–101. doi:10.1007/s13735-019-00190-x

Tran, D. P., & Hoang, V. D. (2019). Adaptive learning based on tracking and ReIdentifying objects using convolutional neural network. *Neural Processing Letters*, *50*(1), 263–282. doi:10.1007/s11063-019-10040-w

Triwiyanto, T., Pawana, I. P. A., & Purnomo, M. H. (2020). An improved performance of deep learning based on convolution neural network to classify the hand motion by evaluating hyper parameter. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *28*(7), 1678–1688. doi:10.1109/TNSRE.2020.2999505 PMID:32634104

Wang, B., Wang, C., Wang, L., Xie, N., & Wei, W. (2019). Recognition of semg hand actions based on cloud adaptive quantum chaos ions motion algorithm optimized SVM. *Journal of Mechanics in Medicine and Biology*, *19*(06), 1950047. doi:10.1142/S0219519419500477

Wang, X. (2018). Research on the improved method of fundamental frequency extraction for music automatic recognition of piano music. *Journal of Intelligent & Fuzzy Systems*, *35*(3), 2777–2783. doi:10.3233/JIFS-169630

Xiao, F., Yang, D., Lv, Z., Guo, X., Liu, Z., & Wang, Y. (2020). Classification of hand movements using variational mode decomposition and composite permutation entropy index with surface electromyogram signals. *Future Generation Computer Systems*, *110*, 1023–1036. doi:10.1016/j.future.2019.11.025

Yan, L. (2019). Design of piano teaching system based on internet of things technology. *Journal of Intelligent & Fuzzy Systems*, *37*(5), 5905–5913. doi:10.3233/JIFS-179172

Zhang, D., & Zakir, A. (2019). Top–Down Saliency Detection Based on Deep-Learned Features. *International Journal of Computational Intelligence and Applications*, *18*(02), 1950009. doi:10.1142/S1469026819500093

Zhang, Z., Zhan, S., Zhang, H., & Li, X. (2020). Joint model of entity recognition and relation extraction based on artificial neural network. *Journal of Ambient Intelligence and Humanized Computing*, 1–9.

Zhi Qian is with the Conservatory of Music, Baoji University of Arts and Sciences, Baoji 721013, China