International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 170-180

https://doi.org/10.70517/ijhsa46413

The effect of real-time feedback generation algorithms on performance skill improvement in piano art instruction

Yanyan Li[1,*]
[1] College of Education, Luoyang Culture and Tourism Vocational College, Luoyang, Henan, 471000, China
Corresponding authors: (e-mail: 18538850303@163.com).

Abstract With the development of educational technology, there are problems such as feedback lag, subjective evaluation, and training fragmentation in traditional piano art instruction. This paper explores the influence of real-time feedback generation algorithm on the improvement of playing skills in piano art instruction. The study designs and implements a deep learning-based piano hand fingering recognition system, which uses the YOLOv3 target detection algorithm to identify playing errors, combines with the HRNet network for gesture estimation, and compares and scores the player's playing with the standard data through the DTW algorithm. Evaluation on the FHAD dataset shows that the average joint error of this system is 16.18 mm, which is better than the comparison methods NTIS of 16.88 mm, Crazyhand of 22.03 mm and BT of 27.24 mm.When the error threshold is above 26 mm, the estimation effect of this system exceeds that of all comparison methods, which indicates that it has an excellent performance in the piano hand fingering feature extraction and fusion with excellent performance. Experiments on two piano art classes (100 students in total) in the College of Music and Arts showed that the average performance of the students in the experimental class that utilized the system for assisted teaching increased by 9.00 points (from 65.45 to 74.45), while the control class that did not utilize the system only increased by 0.79 points, which is a significant difference (P=0.041<0.05). In addition, teaching based on the piano hand fingering recognition system significantly increased students' interest in learning, with the experimental class increasing its interest score from 38.54 to 48.49. The study confirms that the deep learning-based piano hand shape fingering recognition system can effectively improve students' piano playing skills, stimulate learning interest, and provide a new teaching aid for piano art instruction.

Index Terms Piano art instruction, real-time feedback, deep learning, hand fingering recognition, YOLOv3 algorithm, playing skill

I. Introduction

In today's society, the importance of art education is becoming more and more prominent, especially piano art instruction as an important part of music education, which plays an irreplaceable role in cultivating musical talents and improving artistic literacy. Piano art instruction not only requires teachers to have solid piano playing skills, but also needs profound knowledge of music theory and rich teaching experience to guide students to correctly understand and express musical works [1], [2]. With the development of music education and the growing social demand for musical talents, the teaching mode and strategy of piano art instruction also need to be constantly innovated and developed to adapt to the changes of the times [3], [4].

The current teaching mode of piano art instruction is characterized by diversification and complexity, reflecting the adaptation of the field of music education to the ever-changing social needs and technological development. However, in the traditional teaching mode, one-on-one teacher-student face-to-face instruction is still the most common form of piano art instruction, which facilitates teachers to personalize their instruction for students' specific situations, but at the same time, there are problems such as lagging feedback mechanism, strong subjectivity in evaluation, and fragmentation of training, which affects the overall piano playing effect [5]-[8]. These problems highlight the importance of real-time feedback in piano teaching.

With the development of educational technology, digital teaching tools and network teaching platforms began to be widely used in piano art instruction, these new teaching tools not only improve the teaching efficiency and interactivity, but also broaden the access to teaching resources and personalized teaching guidance, enabling students to access a more diverse range of music teaching methods, promoting the students' independent learning ability, and providing real-time feedback on the teaching of the path [9]-[12]. Teaching real-time feedback as a product of educational technology, the use of real-time feedback technology in the piano instruction classroom or after class, through the gesture capture technology, audio recognition technology and other digital technologies to

capture the changes of each note and the details of the student's performance, so that the student can correct their own mistakes in a timely manner, but also to guide the student to think about how to improve, so that the student can continue to adjust their own way of playing in practice, which promotes the independent learning ability, thus gradually improving skills [13]-[16].

In today's society, the importance of art education is becoming more and more prominent, especially piano art instruction as an important part of music education, which plays an irreplaceable role in cultivating musical talents and improving artistic literacy. Piano art instruction not only requires teachers to have solid piano playing skills, but also needs profound knowledge of music theory and rich teaching experience in order to guide students to correctly understand and express musical works. With the development of music education and the growing demand for musical talents in society, the teaching mode and strategy of piano art instruction also need to be constantly innovated and developed to adapt to the changes of the times.

The current teaching mode of piano art instruction is characterized by diversification and complexity, reflecting the adaptation of the field of music education to the ever-changing social needs and technological development. However, in the traditional teaching mode, one-on-one teacher-student face-to-face instruction is still the most common form of piano art instruction, which facilitates teachers to provide personalized instruction for students' specific situations, but at the same time, there are also problems such as lagging feedback mechanism, strong subjectivity in evaluation, and fragmentation of training, which affects the overall piano playing effect. These problems highlight the importance of real-time feedback in piano teaching.

With the development of educational technology, digital teaching tools and network teaching platforms began to be widely used in piano art instruction, these new teaching tools not only improve the teaching efficiency and interactivity, but also broaden the access to teaching resources and personalized teaching guidance, so that the students can be exposed to a more diversified approach to music teaching and promote the independent learning ability of the students, and provide real-time feedback for the teaching of the Path. Teaching real-time feedback as a product of educational technology, the use of real-time feedback technology in the piano instruction classroom or after class, through gesture capture technology, audio recognition technology and other digital technologies to capture every note change and details of the student's performance, so that the student can correct their own mistakes in a timely manner, but also to guide the student to think about how to improve, so that the student can continue to adjust their own way of playing in practice, which promotes the independent learning ability, thus gradually improving their skills.

Based on the above background, this study designs and implements a deep learning-based piano hand fingering recognition system to improve piano playing skills through real-time feedback generation algorithms. The system mainly consists of a playing error recognition module, a hand shape recognition scoring module, and a historical data module, which uses the YOLOv3 target detection algorithm to recognize playing errors, combines with the HRNet network for hand gesture estimation, and compares and scores the player's playing with the standard data by the Dynamic Time Warping (DTW) algorithm. The study experimentally verifies the recognition accuracy of the system on different types of piano music and different piano hand fingerings, and evaluates the actual effect of the system on the improvement of students' piano playing skills through teaching experiments. The results of this study are expected to provide new teaching aids for piano art instruction and promote the innovative development of piano teaching mode.

II. Deep learning based piano hand fingering recognition system

II. A. Overall system architecture

The system is composed of a front-end and a back-end, with the back-end mainly realizing video reading and algorithm recognition, and the front-end displaying the results of the processing through web pages. Firstly, the system acquires the corresponding piano playing video in real time through the camera, after image processing, and then sends it to the algorithm recognition module for piano playing error recognition, hand shape error recognition and scoring. Finally, the recognized result video is displayed in the front-end through WebSocket communication. The algorithm recognition module counts the number of playing errors and the number of hand shape errors, and displays the errors and the overall score on the front-end via HTTP communication. Algorithm recognition is composed of data acquisition, data labeling, model training, algorithm deployment and other steps. Data acquisition is mainly accomplished by collecting videos of various playing error hand shapes and wrong notes during piano playing. Model training is performed using YOLOv3 target detection algorithm, and algorithm deployment is realized by integrating the trained model into the system and deploying it on the server, and the general framework of the system is shown in Figure 1.

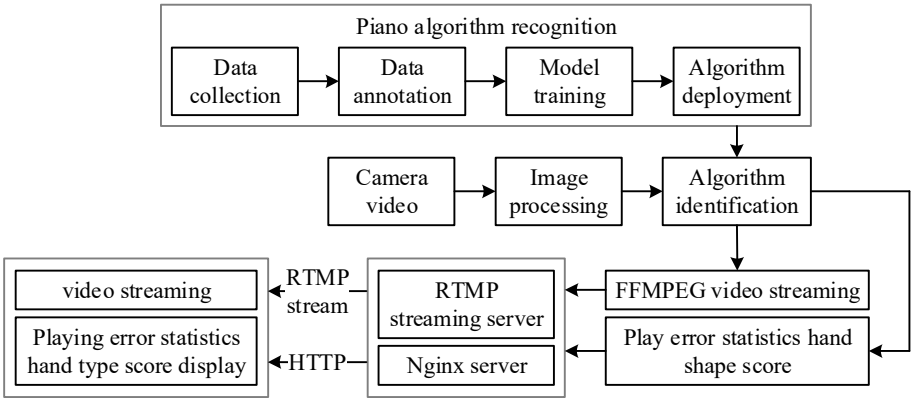

Figure 1: Overall framework of the system

II. B.System Function Module Design

The deep learning-based piano hand shape fingering recognition system designed in this paper realizes the recognition and correction of the practitioner's playing fingering. Through the AI technology simulation accompaniment, guiding students' piano practice, through image and video recognition algorithms, accurately detecting the piano playing hand shape, real-time correction of students' hand shape and fingering errors, mainly realizes the functions of image acquisition, gesture recognition, recognition and comparison, and so on.

The system composition module is shown in Figure 2, the deep learning based piano hand shape fingering recognition system can be divided into playing error recognition module, hand shape recognition scoring module and historical data module, and different functional modules can also be specifically divided into different functional sub-modules.

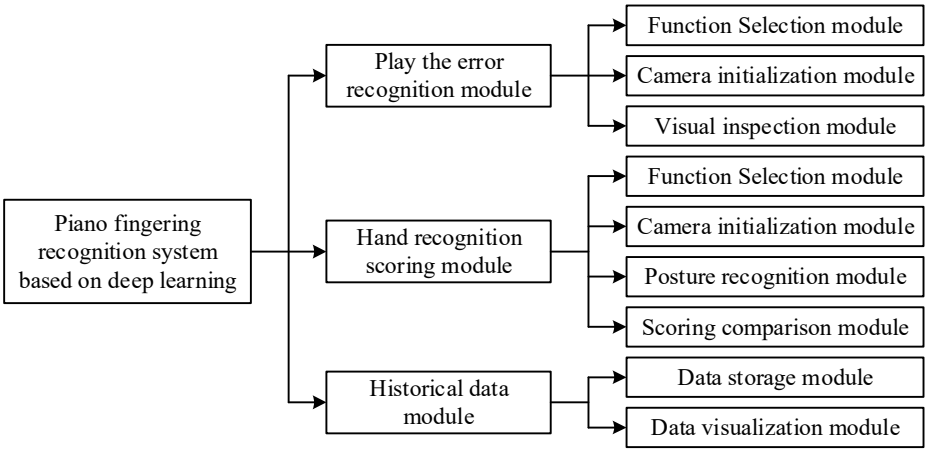

Figure 2: System Composition modules

The specific functions are as follows:

1) Playing error recognition module. This module is mainly used for the practitioner to choose to upload the recording screen detection or call the camera real-time detection, call the target detection algorithm to achieve the recognition of the wrong playing key, real-time will press the wrong key finger in the image screen detection and use the box to mark out, at the same time, the detection results will be real-time pushed to the front-end for display, the practitioner can real-time access to their own play the accuracy of the play.

2) Hand Recognition Scoring Module. The module uses deep learning gesture recognition master method to detect and recognize the playing finger joints, and then compares it with the standard master gesture to output the corresponding score.

3) Historical data module. This module is used to record the user's playing data each time, record the location of each playing error, and output the previously recognized and marked playing data after playing, so that the user can easily see his or her own error methods and technical weaknesses.

II. C.Front-end and back-end video communication

The backend of this system uses the SpringBoot framework, which is an open source application framework on the Java platform with containers with control inversion features. In this paper, we use this back-end framework for web page rendering, routing responses, and database management. Piano playing error recognition and hand shape error recognition algorithms are written in Python language using the deep learning framework PyTorch.

Camera video reading through the Python language Opencv module function cv2.VideoCapture on the practitioner playing real-time image acquisition, and then through the deep learning algorithms for processing and recognition, to get the processing of the frame image, the results of the processing will be converted to Base64 format, the results of the real-time display in the front-end through the WebSocket communication The results are converted to Base64 format and displayed in the front-end in real time through WebSocket communication.

The front-end of the system is designed using the Vue framework, which is a set of progressive frameworks developed based on JavaScript and used to build user interfaces. In this system, we receive the Base64 format data pushed by the back-end WebStock communication in real time through the characteristics of Vue data bidirectional binding, and define the HTML image tag img in the front-end, and render the Base64 format data is rendered and displayed.

III. The main algorithm design of the piano hand fingering recognition system

A series of target detection algorithms and pose estimation algorithms based on deep learning algorithms have made great breakthroughs. The system uses the YOLOv3 algorithm for target detection, the target detection algorithm can realize the detection of the hand position in addition to the recognition of the wrongly played keys, and obtain the hand coordinates of the practitioner in the video, and then the gesture prediction frame picture recognized by YOLOv3 is passed to the gesture estimation algorithm HRnet backbone neural network, and then convolutional and inverse convolutional modules are used to generate the multiple resolution and high resolution solo thermograms to perform the prediction of the gesture recognition joint points.

III. A. Algorithm Architecture Design

The key algorithm for hand recognition in this project consists of two core parts, which are the target detection algorithm and the pose estimation algorithm. The target detection algorithm mainly uses the YOLOv3 algorithm [17] to detect the position where the hand is located in the key frame.

The YOLOv3 algorithm performs target detection by dividing the image into multiple regions and predicting the probability and bounding box of each region. The algorithm is divided into two main steps: first, a series of candidate regions are generated on the image based on certain rules, and then these candidate regions are labeled by their positional relationship with the real frame.

Second, image features are extracted using a convolutional neural network to predict the locations and categories of the candidate regions. Each prediction frame is considered as a sample, and the labeled values are generated by labeling the positions and categories of the real frames relative to the prediction frames, and the predictions of positions and categories are made by the network model.

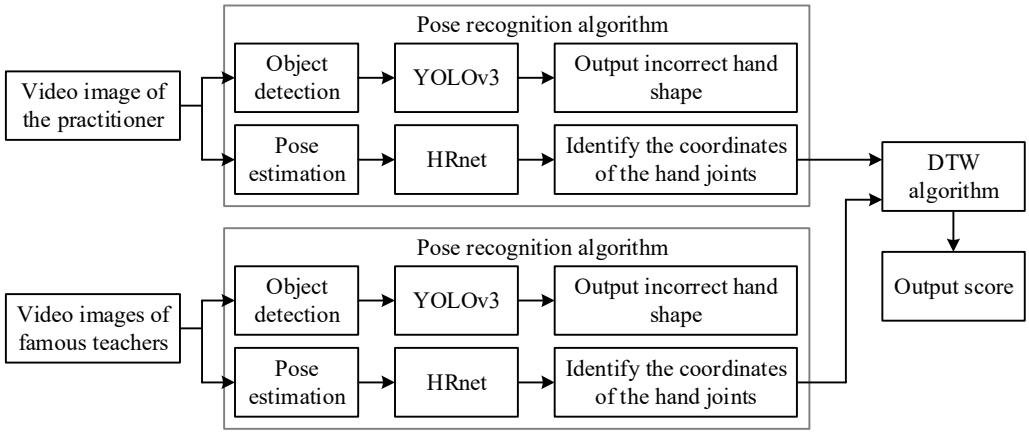

Figure 3: Block diagram of the scoring algorithm

Then two functions are extended on this basis, the first function is to make an error hand shape judgment on the hand shape of the current key frame by YOLOv3 algorithm, and the second function is to predict the hand key point

by using the “HRnet” pose estimation algorithm after obtaining the coordinates of the hand position of the current key frame, and then the coordinates of hand key point are obtained by using “HRnet” pose estimation algorithm. The obtained coordinates of the hand keypoints are stored in an array, and the DTW algorithm is used to compare the player's playing situation with the standard data played by the piano teacher and score the player's performance. The final output is the number of incorrect hand shapes and the score of the player's performance. The block diagram of the scoring algorithm of this system is shown in Fig. 3.

III. B. Target Detection Algorithm

The YOLOv3 algorithm uses the Darknet-53 structure as the backbone network to extract features, employs the cross-entropy loss function to realize the classification task with multiple labels, and uses multi-scale training to improve small target detection. The basic principle is to input a fixed-size image into the network and use the regression idea to obtain the location of the bounding box and the category it belongs to.

The Darknet-53 network is equipped with 52 convolutional layers and one fully connected layer, and features are extracted alternately using convolutional kernels of 1×1 and 3×3 sizes. To prevent the network from deepening and causing gradient dispersion, shortcut connections are added between convolutions to incorporate residual ideas. The image is divided into $N \times N$ grids, and three detection layers are designed for multiscale training after the fully connected layer. Each detection layer predicts the location information, confidence level, and target category of the target bounding box respectively. A single image passing through the detection layer outputs the offset of the bounding box, the target confidence score, and the number of predicted categories.

The upsample up-sampling operation in the network acquires the features of the previous detection layer and combines the shallow network features with the current network features through the concat cascade operation, thus fusing the rich semantic features in the deep network and the basic features in the shallow network.

YOLOv3 regresses the bounding box by k-means clustering nine fixed-size a priori box anchorboxes, assigning three larger-size anchorboxes to the y_1 layer, three smaller-size ones to the y_3 layer, and the rest to the y_2 layer. The algorithm produces multiple candidate frames for the same target that exist overlapping with each other, and the best target frame bounding box is found by the non-great value suppression algorithm, i.e., the target frame bounding box with the highest confidence is obtained after eliminating redundancy.

YOLOv3 loss function is divided into bounding box position regression loss, target confidence loss, target classification loss. Among them, the bounding box position regression loss is divided into center point coordinates and width-height offset loss.

$$L_{xy} = (2 - wh)\sum_{i=0}^{N^2}\sum_{j=0}^{B} 1_{ij}^{obj} \begin{Bmatrix} -t_{xy}\log(\sigma(p_{xy})) \\ -(1-t_{xy})\log(1-\sigma(p_{xy})) \end{Bmatrix} \tag{1}$$

$$L_{wh} = (2 - wh)\sum_{i=0}^{N^2}\sum_{j=0}^{B} 1_{ij}^{obj}(t_{wh} - p_{wh})^2 \tag{2}$$

$$L_{cls} = \sum_{i=0}^{N^2} 1_i^{obj}\left\{-t_{ck}\log(p_i(c)) - (1-t_{cls})\log(1-p_i(c))\right\} \tag{3}$$

$$L_{conf} = -\sum_{i=0}^{N^2}\sum_{j=0}^{B}\left[\log(C_i)\right] - \lambda_{noobj}\sum_{i=0}^{N^2}\sum_{j=0}^{B}\left[\log(1-C_i)\right] \tag{4}$$

where N^2 is the grid size, each grid predicts B prediction frames, and 1_{ij}^{obj} is the presence or absence of the target to be detected in the j bounding box of the i th grid. t_{wh}, t_{xy} are the true values. p_{wh}, p_{xy} is the network predicted value. $(2-wh)$ is the trade-off objective scale parameter. t_{cls} is the category confidence. $p_i(c)$ is the predicted category probability.

III. C. HRNet network

HRNet [18] was originally designed for human posture estimation task here it is migrated to be used in the task of detecting key points for piano hand fingering recognition.HRNet allows the network to keep outputting high resolution feature maps all the time.Figure 4 shows the network structure of HRNet.

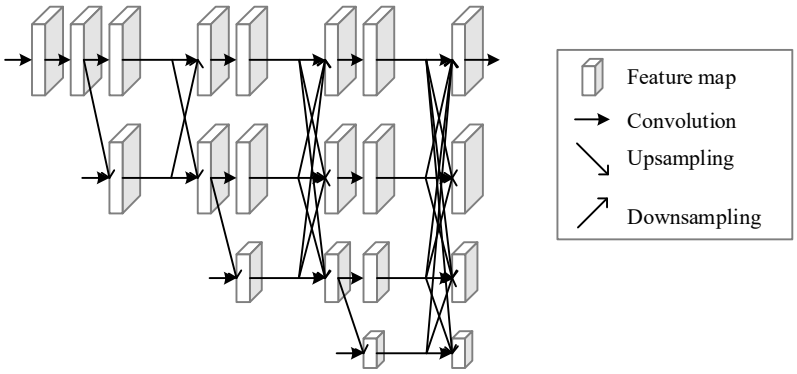

Figure 4: HRNet Network structure

HRNet replaces the way of using low-level features to reproduce high-level information with the way of guaranteeing high-resolution features of the original image, which reduces the loss of spatial accuracy to a certain extent. The exchange unit is added to the multi-resolution subnetwork, which allows repetitive information interaction between feature maps of different scales. For the key point detection of cervical spine X-ray images, the problems of blurred hand structure and low recognition accuracy caused by blurred and low resolution X-ray images are all better solved by HRNet, so this network is chosen as the backbone network.

III. D. DTW algorithm

When different people or even the same person makes the same action, the time of completion and the amplitude of the action will have some differences, which leads to the length of the action sequence generated during data collection may be different from the length of the set standard action sequence, and it is not possible to directly calculate the Euclidean distance between two sequences when calculating the similarity of two actions. In contrast, DTW [19] is an algorithm that can compute the similarity between two time sequences of different lengths, which can adjust the length of the action sequence to be detected through dynamic planning, and subsequently realize the similarity assessment of the two action sequences by calculating the cumulative shortest distance to the template piano-manual action sequence. The calculation using only the Euclidean distance obviously causes great errors.

W to t denote the alignment or mapping of the time series $x(i)$ and $y(j)$, $W=\{w_1(i,j),w_2(i,j),\cdots,w_k(i,j)\},k=1,2,\cdots,p$, $p\in[\max(m,n),m+n-1)$, and p denotes the length of W. In this case, the regularized path has three constraints:

(1) Boundary condition: the starting point of W must be $w_1(1,1)$ and the ending point must be $w_k(m,n)$.

(2) Continuity: Let the neighboring elements in W be $w_{k-1}(i',j')$ and $w_k(i,j)$, then the $i-i'\le 1$ and $j-j'\le 1$ are to be satisfied conditions.

(3) Monotonicity: Let the neighboring elements in W be $w_{k-1}(i',j')$ and $w_k(i,j)$, then the conditions of $i-i'\ge 0$ and $j-j'\ge 0$ should be satisfied. conditions.

The DTW algorithm constructs a $m\times n$ grid matrix with $x(i)$ sequence lengths m as rows and $y(i)$ sequence lengths n as columns, and under the constraints of regularized paths W, the DTW algorithm plans the optimal starting-point-to-terminal paths by searching for the grid points with the shortest cumulative distances $\gamma(i,j)$ obtained by accumulating the $d(i,j)$. Path. The cumulative distance $\gamma(i,j)$ of any point in the grid is calculated in equation (5):

$$\gamma(i,j)=\begin{cases} d(i,j) & i=1,j=1 \\ \gamma(i-1,j)+d(i,j) & i>1,j=1 \\ \gamma(i,j-1)+d(i,j) & i=1,j>1 \\ \min(\gamma(i-1,j),\gamma(i,j-1),\gamma(i-1,j-1))+d(i,j) & i>1,j>1 \end{cases} \tag{5}$$

where $d(i,j)$ is the Euclidean distance between elements in the sequence $x(i)$ and elements in the sequence $y(i)$, $\gamma(i,j)$ is the cumulative distance of the grid where $d(i,j)$ is computed iteratively from (1, 1) to (i,j), and

$\min(\gamma(i-1,j),\gamma(i,j-1),\gamma(i-1,j-1))$ indicates that the point with the smallest cumulative distance is selected to continue the iterative calculation.

In this paper, the piano hand fingering action feature matrix F_m to be measured replaces $x(i)$, and the template action feature matrix $F_m^{'}$ replaces $y(j)$, and the computation of $d(i,j)$ is shown in Eqn. (6), since the feature matrix is composed of a combination of angle features and distance features:

$$d(i,j)=\sqrt{\sum_{k=1}^{l}(F_m(i,k)-F_m^{'}(j,k))^2} \tag{6}$$

where $F_m(i,j)$ is the k th eigenvalue of the i th frame eigenvector in the action feature matrix F_m, $F^{'}m(i,j)$ is the k th eigenvalue of the j th frame eigenvector in the action eigenmatrix F_m, and l is the number of eigenvalues.

The gesture recognition algorithm is used to detect the hand joints when playing the piano to get the hand shape detection effect, and finally the DTW algorithm is used to realize the comparison with the hand shape of the famous teacher to get the score.

IV. Piano hand fingering recognition experiment and result analysis

In order to verify the network of this paper for different types of piano music and different piano hand fingering, the FHAD dataset that meets the requirements was selected for the experiments. In the experiments, the network performance was comprehensively tested according to the pre-set division of the dataset, and compared and analyzed with other methods. Through this series of experiments, the stability and accuracy of the system in this paper in complex scenarios are verified, and a more reliable and effective solution is provided for the task of piano hand fingering recognition.

IV. A. Data pre-processing

Data preprocessing was performed on the input image using image alignment and image cropping. In order for the input data to contain more information about the piano hand fingering features and to weaken the influence of irrelevant interfering information on the results of the piano hand fingering gesture estimation, the image cropping operation is required. In this experiment, the 2D coordinates of the MCP joint of the middle finger are considered as the center of the hand region for image cropping to improve the network's focus on the piano hand fingering focus information.

TRAINING: Initialize the network weights by setting the mean to 1 and the standard deviation to 0.02, and set the weights randomly in the form of a normal distribution. When training the network, the initial learning rate is set to 0.002. As the training proceeds, the learning rate will be dynamically adjusted, and every 200 rounds of arithmetic, the learning rate will be reduced to 0.5 times of the original, and the entire training process of the network is 5000 rounds. In this paper, the network model is implemented under the PyTorch framework, and the whole experiment is done on RTX 2070 GPU and Intel i7-10750H CPU to complete the model training and testing tasks.

IV. B. Assessment of indicators

In order to comprehensively evaluate the performance of the network in this paper for 3D piano hand fingering recognition, three widely used metrics are adopted: mean joint error (MJE), and percentage of correct key points (PCK). Specifically, mean joint error (MJE) is used to quantify the Euclidean distance in millimeters (mm) between the estimated 3D joint coordinates and the true coordinates labeled in the dataset, and this metric intuitively responds to the network's accuracy in predicting the joint positions. Percentage of Correct Keypoints (PCK) describes what percentage of the coordinates of the predicted joints fall within a given range of labeled true coordinates and a given distance. By considering these two metrics together, we are able to comprehensively evaluate the performance of the network in the piano hand fingering recognition task.

IV. C. Comparison of experimental results

On the FHAD dataset, Figure 5 shows the experimental results of this system compared with other methods. The results of the experiments by piano song category are shown in Table 1, where the comparison methods are all from the Open Challenge HANDS 2019, which categorizes the piano song category into classical, pop, and ethnic. Among them, the comparison algorithm 1: NTIS method, which is a method based on V2V network, takes the depth image as the input data, converts the depth image into voxels, and makes a prediction for each voxel. Comparison Algorithm 2: CrazyHand method, which employs a tree branching structure to independently predict the hand shape for each finger of the hand. Comparison Algorithm 3: The BT method, which predicts the fingers separately from the

palm of the hand, processes the point cloud data by voting in order to obtain the labels of the fingers and the palm of the hand, and predicts the results by regression. The average joint point error value of the system in this paper is 16.18mm, which is reduced by 0.70mm compared to 16.88mm of the NTIS method, and is significantly better than the other compared methods. Fig. 5 shows the comparison of the PCK curves of the algorithms, and the estimation of this paper's network outperforms all other methods when the error threshold is above 26mm, which indicates that this paper's network performs very well in piano hand fingering feature extraction and fusion in the face of severe occlusion.

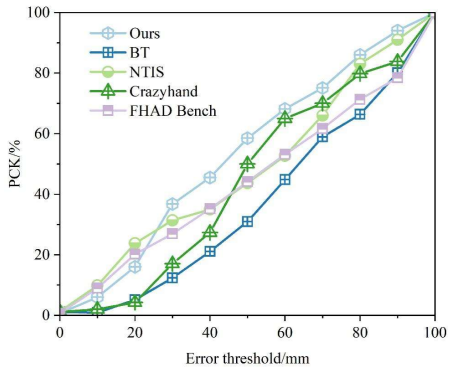

Figure 5: The PCK curve of different methods when divided by the type of music

Table 1: The average error of the category of music

Method	The error of the node of classical music	The customs node error of popular music	The error of the national music	Average node error (mm)
Ours	16.45	17.85	14.25	16.18
NTIS	16.98	18.44	15.23	16.88
Crazyhand	18.54	20.56	26.98	22.03
BT	24.11	28.51	29.11	27.24
FHAD Bench	20.56	22.56	19.24	20.79

After dividing the data by the seven fingering action types of piano such as vertical key down, horizontal movement, legato, staccato, skip, arpeggio, vibrato, etc., the results of algorithm comparison are shown in Table 2 and Fig. 6, which compares the spatial mesh prediction method as well as the benchmark method that comes with the FHAD dataset this time. The proposed method in this paper shows good estimation both in terms of average joint point error and PCK, and the average joint point error of the network in this paper is as low as 11.05 mm.The method in this paper exhibits a good estimation ability in different poses of object-hand interaction, which predicts the action of each finger, and compared to a network that predicts the whole hand, this approach in hand posture possesses higher flexibility.

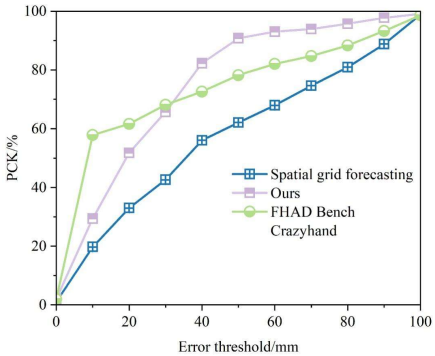

Figure 6: The results and contrast of the PCK curve in the type of piano action

Table 2: The data of the type of action of the piano

Method	Vertical key	Horizontal movement	Play with	Break	Hop	Arpeggio	Seismic tone	Average node error (mm)
Spatial grid forecasting	15.89	16.35	17.56	14.23	13.89	16.54	17.89	16.05
FHAD Benchmark	14.56	15.89	16.54	14.89	15.12	16.99	18.12	16.02
Ours	11.44	12.56	10.23	12.65	9.56	8.98	11.89	11.05

The visualized part of the piano hand fingering recognition results processed by the system in this paper is shown in Fig. 7, in which the true labeled positions of the hand joints are marked by the red line, and the estimated coordinates of the system in this paper are marked by the purple line, and the blue ones are the occluding objects. From the visualization result figure, it can be found that the piano hand fingering obtained by the network estimation is basically consistent with the true labeling in the dataset, and the system can still estimate the location of the key point relatively accurately in the case of an object occlusion or the limitation of the first viewpoint.

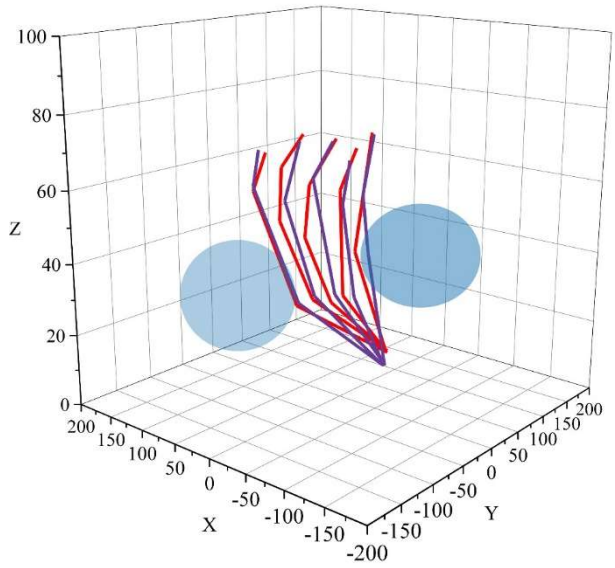

Figure 7: Three-dimensional coordinate prediction on fhad data set

V. Effectiveness of Piano Hand Fingering Recognition System for Improving Playing Skills

V. A. Subjects and experimental design

The research object of this paper is two piano art classes in the sophomore year of a college of music and art, the class size is 50 students, and the two classes have a total of 100 students, named after the experimental class and the control class, which includes: the age, male to female ratio, enrollment scores and other inquiries, and the difference between the piano playing level of experimental class and control class is not very obvious.

The piano playing level is tested through monthly examinations to understand the students' mastery of basic piano playing skills. The experimental class, i.e. Class A, utilized the Piano Hand Shape Fingerings Recognition System to support teaching. The control class, i.e. Class B, was not taught with the Piano Hand Pattern Recognition System.

V. B. Results of testing the effect of improved performance skills

Before the implementation of the experiment, the situation of the students' piano playing level was tested, the total score of the test paper was 100 points, the coefficient of difficulty was 0.740, the test time was 30 minutes, and the statistical results of the test are shown in Table 3. The data in the table can be seen, in terms of piano playing level, the difference between the experimental class and the control class is not too obvious, the overall quality of piano playing skills is more or less the same, the difference in the significance of the situation P = 0.134> 0.05.

Table 3: Test the performance test of the piano

Experimental object	N(total number of students)	MMM (The average performance of piano art)	Sd(standard deviation)	T test
Laboratory class	50	65.45	11.56	0.19
Cross-reference class	50	64.98	13.56	

After that, the piano playing level of all subjects was tested, and the piano playing test scores of the control group and the experimental group before and after training were compared and analyzed respectively, and the results are shown in Table 4.

After the training, the piano playing performance of the experimental group was improved to a certain extent, while the piano playing performance of the control group was not improved to a large extent, and the difference was relatively obvious, and the difference between the experimental group and the control group in the degree of improvement was also obvious, which proved that the use of the Piano Hand Fingering Recognition System for teaching can make the students' piano playing performance improve significantly, and the difference in the case of the significance of the difference is P=0.041<0.05.

Table 4: The results of the experimental group and the test group were raised

Experimental grouping	N	Pre-training test		Post-training test		Performance improvement	
		Average performance	Standard deviation	Average performance	Standard deviation	Improve	T test
Laboratory class	50	65.45	11.56	74.45	9.89	9.00	2.23*
Cross-reference class	50	64.98	13.56	65.77	11.56	0.79	

V. C. Analysis of Learning Interests of Students in Experimental and Control Classes

Due to the development of the market economy there are some changes in values that affect students' interest in piano art learning. Generally speaking, the influencing factors of piano art learning interest can be divided into two levels: direct and indirect, the former is caused by the learning content and process itself, and the latter is caused in the daily piano art learning activities, and the characteristics of the students' interest in learning, motivation, will, and the level of intellectual development (quality) will have a direct impact on the students' piano art learning effect.

Table 5 shows the comparison of learning interest before and after the experiment between the experimental class and the control class. sig is a significance index, generally greater than 0.05 to reject the original hypothesis, the author's one-way test of the statistics of the experimental group and the comparison group before the experiment yielded a sig value of 0.009, which is less than 0.05, indicating that the experimental group and the comparison group do not have any significant differences in their interest in learning the art of the piano before the experimental training. And the sig value of 0.000 from testing the statistics of the two class groups after the experimental training was found to be 0.000, which accepted the original hypothesis, indicating that the teaching training based on the piano hand fingering recognition system stimulates the interest of ordinary students in learning the art of piano.

Table 5: The experiment was compared with the study of the comparison

Experimental grouping	Laboratory class		Cross-reference class		Sig.(2-tailed)
	Average interest	Standard deviation	Average interest	Standard deviation	
Preexperiment	38.54	7.454	38.44	7.451	0.009
After the experiment	48.49	6.598	38.56	7.558	0.000

VI. Conclusion

In this study, the impact of real-time feedback generation algorithms on the improvement of piano playing skills was investigated by constructing a deep learning-based piano hand fingering recognition system. The study draws the following conclusions:

The piano hand fingering recognition system based on YOLOv3 target detection algorithm and HRNet gesture estimation network shows high recognition accuracy. In the experiments divided by types of piano fingering movements, the average joint error of the system for seven different fingering movements was 11.05 mm, which was significantly lower than that of the spatial grid prediction method (16.05 mm) and the FHAD benchmark method

(16.02 mm). In particular, the best performance was achieved in arpeggio fingering recognition (8.98mm) and skip fingering recognition (9.56mm), which indicates that the system is capable of recognizing fine movements.

In different types of piano music recognition experiments, the system has a joint error of 16.45mm for classical music, 17.85mm for popular music, and 14.25mm for ethnic music, which is overall better than other comparison methods. When the error threshold of PCK curve is above 26mm, the estimation effect of this system exceeds that of all comparison methods, which reflects the excellent performance of the system in the extraction and fusion of piano hand fingering features.

The results of the teaching experiment show that the use of this system for assisted teaching can obviously stimulate students' learning interest. The learning interest score of students in the experimental class increased from 38.54 before the experiment to 48.49 after the experiment, while the learning interest of students in the control class remained almost unchanged (from 38.44 to 38.56). The difference in interest between the two groups after the experiment is significant (P=0.000), indicating that the system can effectively promote students' learning motivation.

In summary, the deep learning-based piano hand fingering recognition system can effectively improve students' piano playing skills and stimulate learning interest by providing real-time feedback, providing a new teaching aid for piano art instruction. This research is of great significance in promoting the innovation of piano teaching mode and improving the quality of piano teaching.

References

[1] Zhang, X., & Daoruang, K. (2024). The Development of Piano Teaching for Piano Pedagogy Course. Journal of Ecohumanism, 3(8), 10782-10802.

[2] Ang, K., Lewis, R., & Odendaal, A. (2025). The meanings of professional development: Perspectives of Malaysian piano teachers. Research Studies in Music Education, 47(1), 109-128.

[3] Yin, X. (2023). Educational innovation of piano teaching course in universities. Education and Information Technologies, 28(9), 11335-11350.

[4] Liu, X., & Phokha, P. (2024). The Development of Integrated Piano Teaching Strategies to Enhance Music Skills for Higher Education Level. Journal of Roi Kaensarn Academi, 9(9), 1293-1310.

[5] Hietanen, L., Enbuska, J., Tuisku, V., Ruokonen, I., & Ruismäki, H. (2018). Student teachers' needs in blended piano studies for clinic style face-to-face guidance. The European Journal of Social & Behavioural Sciences, 23(3), 2701-2712.

[6] Zhou, L. (2021). The experimental research on stylized teaching method used in the art guidance of Chinese undergraduates (piano accompaniment)—take the teaching of haidun piano sonata as an example. Arts Studies and Criticism, 2(4), 726-743.

[7] Hamond, L. F., Welch, G., & Himonides, E. (2019). The pedagogical use of visual feedback for enhancing dynamics in higher education piano learning and performance. Opus, 25(3), 581-601.

[8] Lappe, C., Lappe, M., & Keller, P. E. (2018). The influence of pitch feedback on learning of motor-timing and sequencing: A piano study with novices. PLoS One, 13(11), e0207462.

[9] Yang, Z. Y. (2020). Modern piano teaching technologies: Accessibility, effectiveness, the need for pedagogues. Ilkogretim Online, 19(3).

[10] Ruan, W. (2024). Increasing student motivation to learn the piano using modern digital technologies: independent piano learning with the soft Mozart app. Current Psychology, 1-11.

[11] Zheng, Y., & Wang, L. (2024). Application of entertainment virtual technology based on network information resources in piano teaching. Entertainment Computing, 50, 100675.

[12] Sun, J. Q. (2023). Interactive piano Learning Systems: implementing the Suzuki Method in web-based classrooms. Education and Information Technologies, 28(3), 3401-3416.

[13] Hamond, L., Himonides, E., & Welch, G. (2020). The nature of feedback in higher education studio-based piano learning and teaching with the use of digital technology1. Journal of Music Technology & Education, 13(1), 33-56.

[14] Lee, J., Cella, C., & Crayencour, H. C. (2022, September). Vivace: Web Application for Real-Time feedback on Piano Performance. In 3rd Conference on AI Music Creativity.

[15] Wang, Y., Yao, J., & Wang, Z. (2024, December). PianoPal: A Robotic Multimedia System for Interactive Piano Instruction Based on Q-Learning and Real-Time Feedback. In International Conference on Multimedia Modeling (pp. 201-214). Singapore: Springer Nature Singapore.

[16] Zeng, M. (2024). Gesture Recognition Technology of VR Piano Playing Teaching Game based on Hidden Markov Model. International Arab Journal of Information Technology, 21(4), 760-772.

[17] Sadayuki Ito,Hiroaki Nakashima,Naoki Segi,Jun Ouchida,Ippei Yamauchi,Takashi Hirai... & Shiro Imagama. (2025). Development of a YOLOv3-Based Model for Automated Detection of Thoracic Ossification of the Posterior Longitudinal Ligament and the Ligamentum Flavum on Plain Radiographs. Journal of Clinical Medicine,14(7),2389-2389.

[18] Xinghong Yang,Heqing Li,Wei Zhu & Yi Zuo. (2025). RSHRNet: Improved HRNet-based semantic segmentation for UAV rice seedling images in mechanical transplanting quality assessment. Computers and Electronics in Agriculture,234,110273-110273.

[19] Zhonghai Chen & Tengyu Zhang. (2025). Evaluation of basic sports actions for students based on DTW posture matching algorithm. Systems and Soft Computing,7,200196-200196.