Journal of Network Intelligence
Taiwan Ubiquitous Information
 ISSN 2414-8105
Volume **10**, Number **2**, May **2025**

Piano Soundboard Classification Based on Intelligent Neural Network and Multi-feature Fusion Algorithm

Jie-Ru Huang*
Jiangxi Normal University, Nanchang 330027, P. R. China
fabiaolunwen0924@163.com

Mei-Chen Liu
Jiangxi Normal University, Nanchang 330027, P. R. China
lunwenshida1014@163.com

Jin-Min Zhou
Computer Sciences Corporation
Annapolis Junction, MD 20701, United States of America
hk4030@163.com

*Corresponding author: Jie-Ru Huang
Received March 7, 2024, revised August 2, 2024, accepted November 30, 2024.

Abstract. Aiming at the current piano soundboard classification algorithm feature extraction is not sufficient, leading to low classification accuracy, this article suggests a piano soundboard classification algorithm based on intelligent neural network and multi-feature fusion algorithm. Firstly, for the traditional Convolutional Neural Network (CNN) network computation time-consuming issue, adopting the channel attention mechanism to share weights reduces the amount of parameters of the model and enhances the accuracy of the model. Secondly, the grayscale covariance matrix and Principal Component Analysis (PCA) algorithm are adopted to extract the texture features and color features of the soundboard to enhance the texture and color details, respectively. Then the texture features and color features are consistency processed and input to the multi-scale convolutional layer to obtain the local features, and the local features are fused with the depth features by transposed convolutional mapping to achieve the global features. Finally, the achieved consistent features, local features and global features are fed into different channels of the optimized CNN model for classification, and the prediction values of each channel are fused as the final prediction results. The experimental outcome indicates that the suggested method has a high accuracy, precision, recall and F1 value as well as a short classification time, which exhibits a excellent classification performance.
Keywords: Soundboard classification; Convolutional neural network; Multi-feature fusion; Principal component analysis; Channel attention mechanism

1. **Introduction.** Recently, with the continuous enhancement of material living standards, people pay more and more attention to the pursuit of the spiritual world. As a carrier of people's emotions, piano is increasingly favored by consumers, and its quality has been paid attention to [1]. The piano soundboard is one of the core parts of the piano and has a decisive influence on the sound quality [2]. For the goal of ensuring the performance of the soundboard, piano manufacturers need to grade according to the comprehensive surface characteristics such as the texture and color of the soundboard, and assemble the same soundboard to ensure the quality to meet the needs of consumers for

the acoustic quality of the piano. At present, many domestic soundboard wood processing is still on the ground of manual detection, but owing to the difference in the understanding of classification standards and detection level of different inspectors, and long-term mechanical repetitive labor is easy to lead to fatigue of inspectors, with the passage of detection time, the wrong detection rate will gradually increase. It is hard to ensure the stability of wood classification quality [3, 4, 5].

1.1. **Related work.** The texture information of piano soundboard mainly includes two categories: wood grain and vibration mode texture, which have a significant impact on the sound quality of piano. Haralick et al. [6] suggested the famous gray level co-occurrence matrix (GLCM) to describe the texture information of piano soundboard. Walker et al. [7] analyzed the GLCM method from the perspective of genetic algorithm and adaptive multi-scale algorithm, respectively, and reduced the classification error detection rate. Mallat [8] introduced wavelet analysis into the texture research of piano soundboard for the first time, providing an accurate and unified framework for time-frequency multiscale texture analysis. Ojala et al. [9] offered a Local Binary Pattern (LBP) to describe the local texture features of piano soundboards. Gu et al. [10] adopted color vector angle to construct color features of soundboard wood, and used SVM as a classifier to classify colors, but the anti-interference ability was poor. Yoshikawa [11] adopted the color histograms of HSV and LAB color space to extract the color features of soundboard, and used decision trees to classify them. The texture of the soundboard determines how the vibration energy is transmitted on the surface of the soundboard. Wood with different textures can transmit vibration in different ways, thus affecting timbre and volume.

As high-performance Graphics Processing Units (GPU) develop, deep learning has been more widely used owing to its parallelizable computing architecture. Shi et al. [12] extracted the mean color features of soundboard surface and realized the classification of furniture soundboard through BP neural network classifier. Barmpoutis et al. [13] represented the soundboard image as a high-order linear dynamic cascade histogram, and used support vector machine (SVM) classifier to classify the image, but did not consider the internal features of the soundboard. Yang et al. [14] established artificial neural network and Convolutional Neural Network (CNN) respectively to classify five soundboards, and found through comparison that the CNN model had the highest recognition accuracy. Kilic et al. [15] established a CNN neural network model relying on texture data to classify soundboards. Fabijańska et al. [16] suggested the use of convolutional neural networks to automatically identify soundboard categories and determine the final classification through voting. Yang et al. [17] adopted CNN to extract initial features and reduce dimension through principal component analysis to obtain the final texture features, which were input into the SVM for classification, but the classification accuracy was not high.

The soundboard classification method with single feature fails to integrate other features, resulting in low classification efficiency. Hu et al. [18] adopted SVM to extract texture and spectral feature information of soundboard images, and classified them. Han et al. [19] used multi-feature fusion technology to combine gray co-occurrence matrix with various texture features of LBP to enhance the classification accuracy. Wan et al. [20] added a feature fusion module to the traditional neural network model architecture to realize soundboard recognition, but other characteristic information is easily ignored by the model after fusion, resulting in inefficient feature fusion.

1.2. **Contribution.** Focusing on the issue of inadequate feature extraction and low classification accuracy of current piano soundboard classification algorithms, this article designs a piano soundboard classification algorithm based on intelligent neural network and

multi-feature fusion algorithm. The experimental outcome indicates that the accuracy rate, precision rate, recall rate and F1 value of the suggested algorithm are more than 90%, which has a high classification efficiency. (1) The channel attention mechanism is adopted to adapt the model to select the size of one-dimensional convolutional kernel, which reduces the number of parameters in the CNN model and improves the model accuracy. (2) The gray co-occurrence matrix and Principal Component Analysis (PCA) algorithm are adopted to extract the texture and color features of the soundboard to enhance the texture and color details. Then the texture features and color features are processed in a consistent manner and input into the multi-scale convolutional layer to gain local features. The local features are fused with the depth features through transposed convolution mapping to obtain global features. (3) All the extracted features are fed into different channels of the Enhanced CNN (ECNN) model for feature classification, and the weighted average method is adopted to fuse the predicted values of each channel as the final prediction result.

2. Theoretical analysis.

2.1. Convolutional neural network.

CNN adopts a deep nonlinear network framework when learning input image data, which is composed of input level, convolutional level, pooling level, fully connected level and output level [21], as indicated in Figure 1.

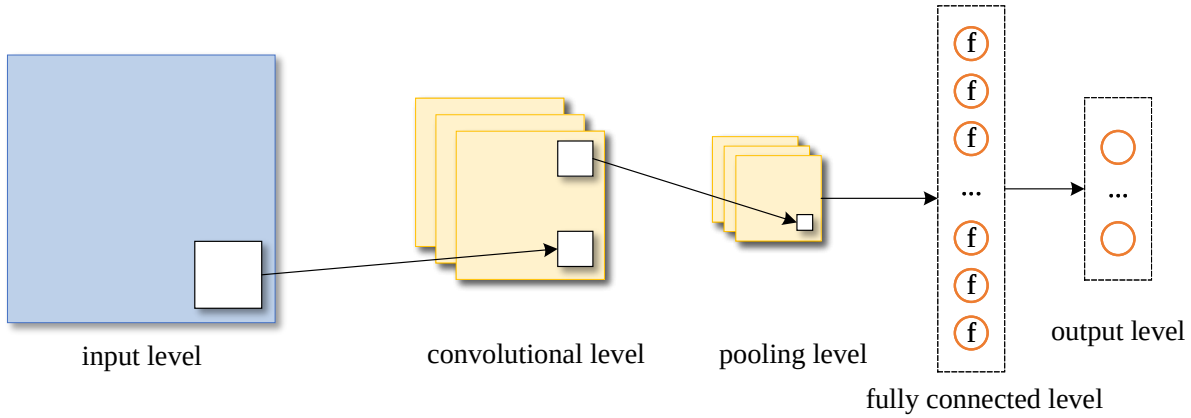

Figure 1. The structure of CNN

(1) Input level. The data input to the model is preprocessed to adapt the data to the model's input.

(2) Convolution level. The convolution kernel slides over the input data in the form of a small matrix, and does the volume operation with the corresponding elements in the data, and finally adds up to an output value, which is calculated as below:

$$f_i^n = \beta\Big(\sum_j h_{ij}^n \cdot f_j^{n-1} + a_i^n\Big) \tag{1}$$

where f_i^n represents the i-th feature of the n level, a_i^n represents the i-th offset value of the n level, and h_{ij}^n represents the convolution kernel of the n level.

(3) Pooling level. Pooling level is usually used after the convolution level to further reduce the data size and extract more significant features after extracting features from the convolution level.

(4) Fully connected level. The fully connected level's function is to vectorize the previous feature map and serve as the final input for predicting classification results or other tasks.

2.2. **Multi-scale feature fusion.** Images contain rich spatial spectral features, and most CNN-based methods can perform the classification task well, but the problems such as the different size of image categories, variable shapes and small targets bring challenges to the realization of high-precision classification. At present, multi-scale feature fusion [22] can address such issues.

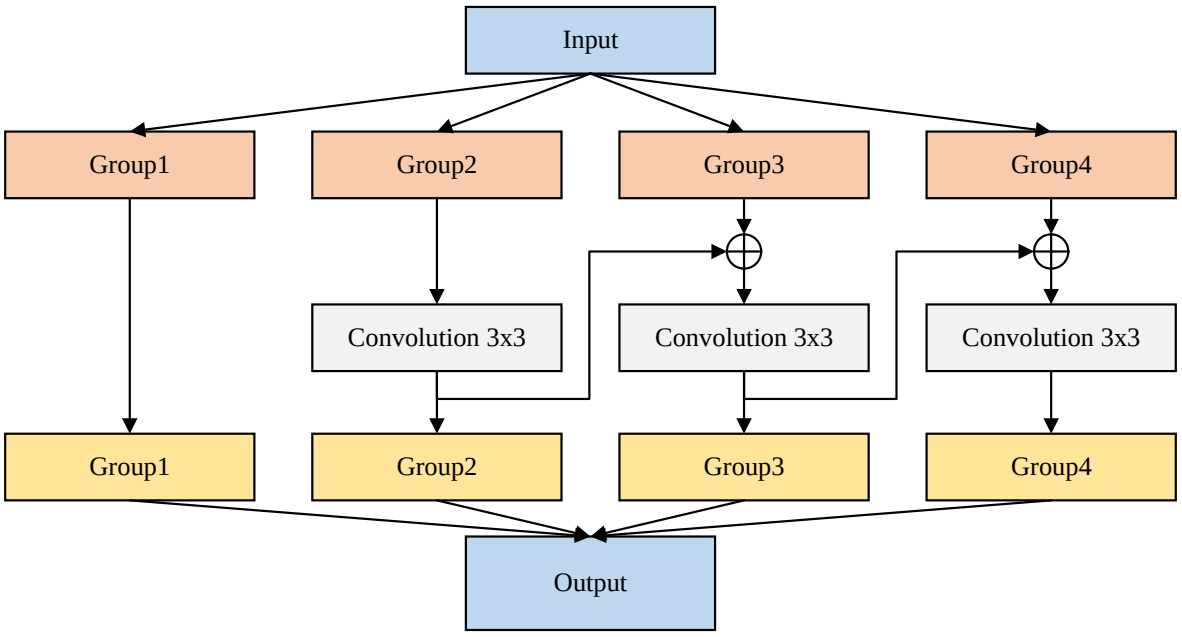

Figure 2. The process of the multi-scale fusion

As indicated in Figure 2, the extraction of these multi-scale features is achieved by designing convolutional or network structures to expand the receptive field of the convolution. For deep features, the receptive field is relatively large and contains stronger semantic information, but the detail perception ability of the obtained feature map is poor. Thus, if the features of different levels in the CNN are effectively fused, the classification performance of the network will be significantly improved.

3. **Optimization of CNN based on channel attention mechanism.** Intending to the issue that traditional CNN networks need to consume a large amount of computational resources to train the samples, this paper utilizes a channel attention mechanism without dimensionality reduction to enhance the CNN (ECNN) model. The learning strategy is used to make the model adaptively choose the size of the one-dimensional convolutional kernel, and the shared weights are used to reduce the number of parameters of the model, which can greatly improve the accuracy of the model with only a small increase in parameters.

The input feature map is divided into two branches and passed backward. In the main branch, Global Average Pooling (GAP) operation is used to compress the feature map and reduce the computation amount of convolution operation. The one-dimensional convolution operation is adopted to extract the features between channels. Input the features into the activation function, process them into the same dimension as the input features, and then multiply them with the corresponding elements of the original features to obtain semantic features with channel information. In this case, the expression for F_{GAP} is as follows:

$$F_{\mathrm{GAP}} = \frac{1}{V \times G} \sum_{i=1}^{V} \sum_{j=1}^{G} X_{ij} \tag{2}$$

where $X_{ij} \in \mathbb{R}^{V \times G \times D}$ denotes the pixel value of a channel in the feature map, and $V \times G$ denotes the spatial size (height $\times$ width), and D is the number of channels.

After the feature map is converted to channel information by GAP, the one-dimensional convolution adopts learning to select the size of the convolution kernel to capture the feature information between channels. The output one-dimensional feature vector is converted to vector R^D containing only channel information, such that $y \in R^D$, v denotes the weights of each channel, and Equation (3) represents the channel attention learning process.

$$\nu_i = \delta\Big(\sum_{j=1}^{l} v_j^i y_j^i\Big), \quad y_j^i \in \Omega_j^l \tag{3}$$

where ν_i represents the information interaction between the i-th channel and its neighboring channels, and then the weight of the current channel is obtained by the nonlinear activation function, and Ω_j^l represents the l channels neighboring the current channel. To simplify the model complexity, reduce the number of parameters, and let all channels share the weight information, Equation (3) can be further simplified to Equation (4).

$$\nu_i = \delta\Big(\sum_{j=1}^{l} v_j^i y_j^i\Big), \quad y_j^i \in \Omega_j^l \tag{4}$$

Relied on the above equation, it can be deduced that the information correlation across channels can be realized by a one-dimensional convolution with a convolution kernel of size l as indicated in Equation (5).

$$\nu = \delta\big(D_l(y)\big) \tag{5}$$

where ν denotes the weight of each channel, and D_l denotes a one-dimensional convolutional operation with convolutional kernel size l. The nonlinear mapping between the channels and the convolution kernel is represented by an exponential function with a base of 2, which is adopted to determine the size of the convolution kernel, as indicated below.

$$\begin{cases} D = \emptyset(l) = 2^{\lambda \times l - a} \\ l = \varphi(D) = \left| \frac{\log_2 D + a}{\lambda} \right|_{odd} \end{cases} \tag{6}$$

where $|(\log_2 D + a)/\lambda|_{\mathrm{odd}}$ turns out to be odd, λ is a hyperparameter, and a is a constant guaranteeing that the one-dimensional convolution kernel size is always odd.

4. **Piano soundboard classification based on intelligent neural network and multi-feature fusion algorithm.**

4.1. **Texture feature extraction for piano soundboard.**

Focusing on the current piano soundboard classification algorithms with large differences in image size and insufficient feature extraction, resulting in low classification accuracy, this article designs a piano soundboard classification algorithm relied on intelligent neural network and multi-feature fusion algorithm. Firstly, the texture features and color features of the soundboard are extracted, and then both of them are consistency processed and input to the multi-scale convolutional level to get the local features, which are fused with the depth features

through transposed convolutional mapping to get the global features. Finally, the obtained coherent features, local features and global features are fed into different channels of the ECNN model for feature classification, and predicted values of each channel are fused as final prediction results using the weighted average method. The whole model is indicated in Figure 3.

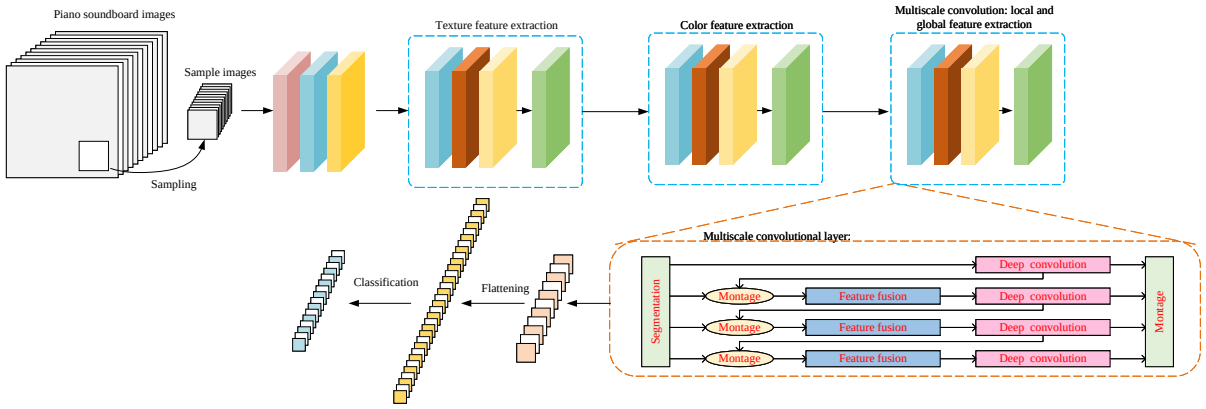

Figure 3. Overall structure of suggested model

This article adopts the gray scale covariance matrix [23] to extract the texture features of piano soundboard, assuming a soundboard image of size $N \times N$, take any point $a : (x, y)$ and another point $b : (x + \Delta x, y + \Delta y)$, which is at a certain distance from it, and set the gray scale value of the pixel pair to be (i, j). Then the features of the soundboard image can be obtained by the two-dimensional spatial features of the image and the convolution operation with the w-th order Laplacian algorithm L, as indicated bellow.

$$f_a^{(w)}(x, y) = \sum_{i=-\frac{N-1}{2}}^{\frac{N-1}{2}} \sum_{j=-\frac{N-1}{2}}^{\frac{N-1}{2}} L^{(w)}(i, j)\, f_b(x + i, y + j) \tag{7}$$

where $f_b(x + i, y + j)$ denotes the gray value of the two-dimensional image of the soundboard image at a distance a band $(\Delta x, \Delta y)$. The convolution operation will exceed the boundaries of $f_b(x + i, y + j)$, so it is necessary to fill the boundaries. Assuming that $f_b(x + i, y + j)$ is the matrix of $Q \times P$, depending on the size of the image, the pixel values of the boundaries of rows $(N - 1)$ are copied on each of the four boundaries of $f_b(x + i, y + j)$ to obtain a new matrix of size $(Q + N - 1) \times (P + N - 1)$. This ensures that $f_b(x + i, y + j)$ can be convolved with the Laplacian operator to obtain the matrix of $Q \times P$ without losing some of the features that lie on the boundaries.

To ensure that the response of the extracted soundboard texture features is on the same scale as the original pixel values, a normalization operation is required before the convolution operation. Finally, the extracted features are summed with the original spatial features to enhance the texture details. For the pixel of the soundboard image in band b, the final w-order feature $F_b^{(w)}(x + \Delta x, y + \Delta y)$ is expressed as bellow.

$$F_b^{(w)}(x+\Delta x, y+\Delta y) = f_b(x+\Delta x, y+\Delta y) + \frac{\sum_{i=-\frac{N-1}{2}}^{\frac{N-1}{2}} \sum_{j=-\frac{N-1}{2}}^{\frac{N-1}{2}} f_b^{(w)}(x + \Delta x, y + \Delta y)}{\sum_{i=-\frac{N-1}{2}}^{\frac{N-1}{2}} \sum_{j=-\frac{N-1}{2}}^{\frac{N-1}{2}} L^{(w)}(i, j)} \tag{8}$$

After extracting the texture features from the original soundboard image using the above equation, assuming that each pixel corresponds to a gray value of s^w, it has the

same M bands as bellow.

$$s^{(w)} = \left[s_0^{(w)}, s_1^{(w)}, s_2^{(w)}, s_3^{(w)}, \ldots, s_{M-1}^{(w)}\right] \tag{9}$$

The final texture feature $\widetilde{X}_T = F \parallel s^{(w)}$ is obtained by concatenating the feature vectors F and $s^{(w)}$.

4.2. Color feature extraction for piano soundboard. In the actual production of the piano, the soundboard texture and color have an important impact on the acoustic quality of the piano, so this paper on the soundboard texture feature extraction should also take into account the soundboard color feature extraction. The PCA algorithm [24] is adopted to extract the color features of the soundboard image.

First, the PCA algorithm is adopted to obtain the color feature subspace of the soundboard image. x_i is the soundboard image, $i = 1, 2, \ldots, m$, m is the total number of images, x_i is used as a training sample, and then the R, G, and B values of the pixel points of x_i are formed into a matrix to compute the sample mean $\text{avg} = \frac{1}{m}\sum_{i=1}^{m} x_i$ of x_i, where avg is the mean value of sample x_i.

Calculate the sample mean $\text{avg}_l = \frac{1}{m_l}\sum_{x_i \in l} x_i$ for each color class separately, where avg_l is the mean of the samples belonging to color class l, m_l is the number of samples belonging to the l-th color class.

The soundboard image samples are normalized as bellow.

$$xx_i = x_i - \text{avg} \tag{10}$$

where xx_i is the normalized soundboard image sample.

The normalized soundboard image samples are formed into a matrix X, and the covariance of this matrix is computed $A = XX^T$, where the upper corner scale T is a normal distribution.

Calculate the color eigenvalues and the eigenvector matrix of A, and arrange the eigenvalues in descending order to build the color eigenvector matrix. The transpose matrix of this eigenvector matrix is the color feature subspace of the soundboard image expressed by $\widetilde{X}_C$. By mapping x_i into the color feature subspace, all the color features of the soundboard image can be achieved as bellow.

$$\widetilde{X}_C = F_C^T X \tag{11}$$

4.3. Multi-scale convolution. After extracting the texture features and color features of the soundboard, the optimized convolutional neural network can effectively characterize the region of interest by learning to emphasize or suppress the relevant information, which helps the network to focus on the relatively important texture features and colors, and enhances the classification performance of the network.

$\widetilde{X}_T$ is the extracted soundboard texture features, $\widetilde{X}_C$ is the extracted soundboard color features, and the pre-trained convolutional neural network is defined as φ_{ECNN}, $\widetilde{X}_T, \widetilde{X}_C \in \mathbb{R}^{l \times m \times n}$ (l and $m \times n$ represent the number and size of the soundboard images, respectively). The consistency algorithm [25] is utilized to concentrate $\widetilde{X}_T$ and $\widetilde{X}_C$ in the high confidence regions, which contain the abundance of important feature information. The consistency index feature $\widetilde{X}_{TC}$ is indicated below.

$$\widetilde{X}_{TC} = 1 - \left|\delta(\widetilde{X}_T) - \delta(\widetilde{X}_C)\right| \tag{12}$$

where 1 represents the all-1 feature vector and $\delta(\cdot)$ denotes the Sigmoid function.

In addition, the local features $\widetilde{X}_{attr}$ achieved by the output feature map after multi-scale convolution through the GELU activation function are indicated below.

$$\widetilde{X}_{attr} = -f^{1\times1}\big(\text{GELU}\big(f^{d\times3\times3}(F_{i-1})\big)\big) + F_{i-1} \tag{13}$$

where F_i denotes the feature map output by the convolutional neural network and $f^{d\times3\times3}$ denotes the convolutional kernel of size 3×3.

The consistency indicator feature $\widetilde{X}_{TC}$ obtained after preprocessing is fed into the ECNN network for dimension expansion so that the dimension of the local feature is the same as that of the deep feature. Secondly, feature $\widetilde{X}_{TC}$ is vector weighted with the local feature $\widetilde{X}_{attr}$ of the soundboard after dimension expansion, so as to use the local feature to control the importance of the deep feature. Final global features achieved are indicated below.

$$\widetilde{X}_{A\text{-}ECNN} = \left[\varphi_{\mathrm{ECNN}}(\widetilde{X}_{attr})\right] \odot \widetilde{X}_{TC} \tag{14}$$

4.4. Multi-feature fusion and classifier construction. To fuse the above multi-feature information, the consistency index feature $\widetilde{X}_{TC}$, local feature $\widetilde{X}_{attr}$ and global feature $\widetilde{X}_{A\text{-}ECNN}$ of the soundboard texture feature $\widetilde{X}_T$ and color feature $\widetilde{X}_C$ are fed to the fully connected layer to obtain the classification prediction values for the three channels of the ECNN.

$$\mathrm{Pred}_{F\text{-}\mathrm{ECNN}_1} = \mathrm{FC}\big(\widetilde{X}_{TC}\big) \tag{15}$$

$$\mathrm{Pred}_{F\text{-}\mathrm{ECNN}_2} = \mathrm{FC}\big(\widetilde{X}_{attr}\big) \tag{16}$$

$$\mathrm{Pred}_{F\text{-}\mathrm{ECNN}_3} = \mathrm{FC}\big(\widetilde{X}_{A\text{-}\mathrm{ECNN}}\big) \tag{17}$$

where $\mathrm{Pred}_{F\text{-}\mathrm{ECNN}_1}$, $\mathrm{Pred}_{F\text{-}\mathrm{ECNN}_2}$ and $\mathrm{Pred}_{F\text{-}\mathrm{ECNN}_3} = \mathrm{FC}(\widetilde{X}_{A\text{-}\mathrm{ECNN}})$ denote the prediction scores of the three channels of the ECNN model for the soundboard image test set, and FC denotes the fully connected layer used for feature classification.

The loss function used for the model is the cross-entropy loss function with the following expression.

$$\mathrm{Loss}_{F\text{-}\mathrm{ECNN}_j} = -\frac{1}{N}\sum_{i=1}^{N} b_i \log\big(\mathrm{Pred}_i^{F\text{-}\mathrm{ECNN}_j}\big) \tag{18}$$

where $j = 1, 2, 3$, $\mathrm{Loss}_{F\text{-}\mathrm{ECNN}_j}$ denotes the loss function of the ECNN model, N refers to the size of the soundboard dataset, b_i refers to the true label of the i-th data, and $\mathrm{Pred}_i^{F\text{-}\mathrm{ECNN}_j}$ refers to the score value of the ECNN model for the i-th data.

Secondly, the predicted values of the three channels of the ECNN model are fused using the information fusion method [26] to improve the classification performance. By assigning different weighting factors to the predicted score values of the ECNN model and calculating the predicted scores after weighted averaging, the fused predicted values are obtained calculated as indicated in Equation (19).

$$P_{\mathrm{fusion}} = \alpha\,\mathrm{Pred}_{F\text{-}\mathrm{ECNN}_1} + \beta\,\mathrm{Pred}_{F\text{-}\mathrm{ECNN}_2} + \gamma\,\mathrm{Pred}_{F\text{-}\mathrm{ECNN}_3} \tag{19}$$

where P_{fusion} is the predictive score generated by the model after feature fusion, and α, β and γ denote the weights assigned to the predictive values $\mathrm{Pred}_{F\text{-}\mathrm{ECNN}_1}$, $\mathrm{Pred}_{F\text{-}\mathrm{ECNN}_2}$ and $\mathrm{Pred}_{F\text{-}\mathrm{ECNN}_3}$, respectively.

5. Performance testing and analysis.

5.1. **Experimental analysis of multi-feature fusion.** To estimate the performance of the suggested the suggested algorithm, the experiments in this article are relied on the modeling of 4182 piano soundboard image datasets randomly selected from an e-commerce platform with five major categories, and the material datasets are divided into the training set, validation set, and test set according to the ratio of 6:2:2. WDBP [12], TSFF [18] and the classification algorithm OURs suggested in this article are trained and tested respectively.

All algorithms are trained on Tensorflow deep learning framework [27] adopting GPU, the input image size is 224×224, the batch size for training iterations is set to 64, momentum is set to 0.9, and the coefficient of the random deactivation dropout function is set to 0.2. The initial value of the learning rate is set to 0.001, and the weights are decayed to 10^{-5}. experimental Platform The configuration is Windows 10, 64-bit operating system, Intel(R) Core (TM) i5-12490F CPU@4.6GHz, 8 GB RAM, and the simulation experiment environment is python v3.0.

This article focuses on the fusion of extracted soundboard color, texture, local and global features and thus the classification of piano soundboard images, with the aim of analyzing whether multi-feature fusion classification is more advantageous compared to single feature. The classification performance metrics are Accuracy, Precision, Recall and the reconciled mean of Precision and Recall F1 [28]. The experimental outcome is indicated in Table 1.

Table 1. Comparison of classification performance for different features

Feature pattern	**Accuracy**	**Precision**	**Recall**	**F1**
color feature	0.6451	0.6109	0.6394	0.6248
texture feature	0.7319	0.7142	0.7426	0.7281
Local feature	0.7628	0.7508	0.7612	0.7560
global feature	0.8137	0.8146	0.8347	0.8245
fusion feature	0.9015	0.8917	0.9135	0.9025

From Table 1, it can be seen that global features have a greater impact on the classification performance than color, texture and local features because global features are obtained by combining deep features extracted adopting convolutional networks with local features, which extracts deeper features of the soundboard image, and thus improves the classification performance compared to the other three single features. The classification performance of the soundboard integrated with color, texture, local and global features is the best, and the Accuracy is improved by 39.75%, 23.17%, 18.18% and 10.79%, respectively, compared with the single feature of color, texture, local and global classification algorithms. Precision has increased 45.96%, 24.85%, 18.77% and 9.46%, Recall has increased 42.87%, 23.01%, 20.11% and 9.44%, and F1 has increased 44.45%, 23.95%, 19.38% and 9.45%, respectively. It indicates that the fusion of features can better characterize the image as a whole and achieve the goal of improving the classification accuracy of soundboard images.

The time-consuming classification of soundboard images with the four single features and multi-feature fusion is indicated in Figure 4. The performance of soundboard image classification using single features is relatively poor, the average macro-F1 value is lower than 0.8. The fusion feature has the best performance, the average F1 value reaches 90.25%, and the classification time is the lowest, which is 38.57 us, and the performance of the color feature is the worst, the average F1 value is 62.48%, and the classification time is 65.12 us, which

Figure 4 indicates that the classification of soundboard image fusion can effectively reduce the computational consumption and improve the classification efficiency.

5.2. **Comparative experimental analysis.** Table 2 gives a comparison of the classification performance of OURs and the other algorithms. From the table, it can be seen that the classification performance of OURs algorithm is better than that of WDBP classification and TSFF classification, which further illustrates the effectiveness of the optimization of CNN. WDBP algorithm, although the use of BP neural network and color features in

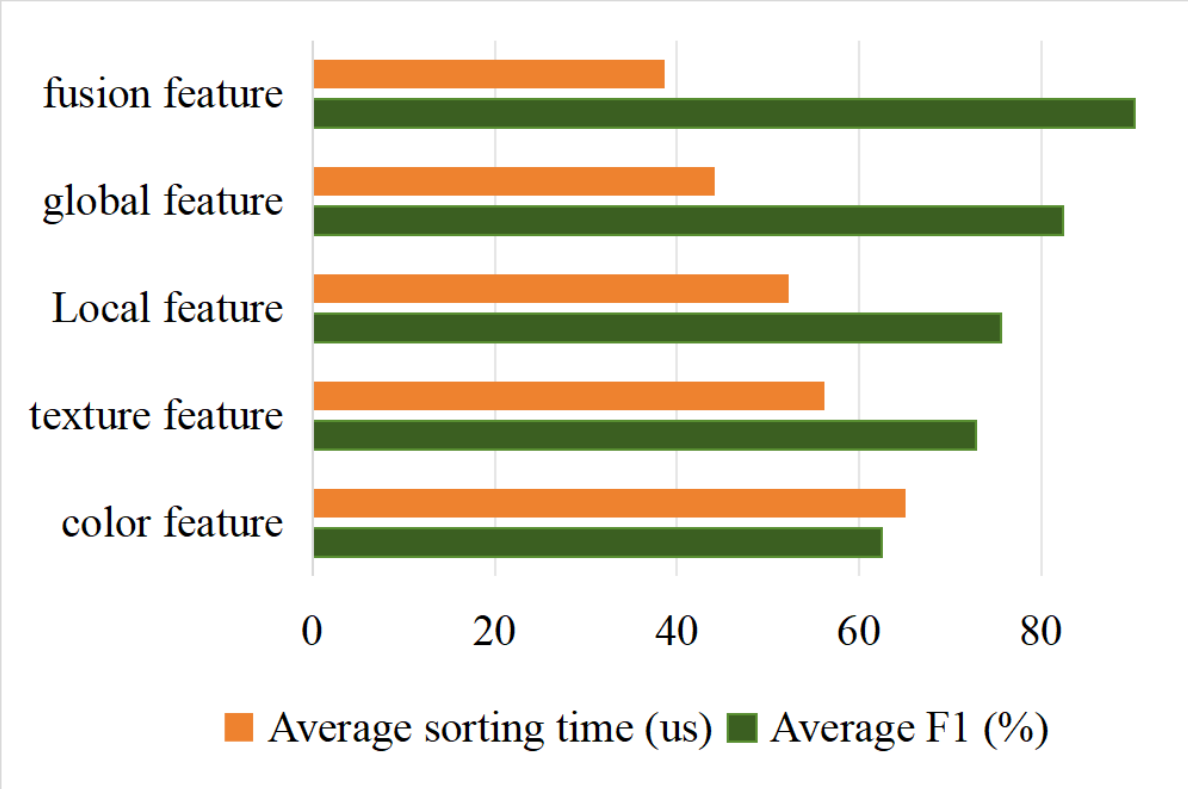

Figure 4. Comparison of time consuming and average F1 for image classification with different features

the classification of soundboard images, did not consider a variety of features of the image, there are the disadvantages of slow transmission of learning, it is easy to fall into the local minimum, making the classification of the worst effect, while the TSFF algorithm is constructed relied on CNN and multiple texture features to classify soundboard images, but did not improve the CNN, and did not consider other attributes of the features, so the classification effect is not good. The OURs algorithm not only optimizes the convolutional neural network, but also takes into account the effective fusion of multiple features in the process of feature extraction, and uses different sizes of convolutional kernels to extract the information weights conducive to classification in a multi-scale manner so that the network reduces the transmission of noise information as much as possible, thus improving the classification performance of the whole network model.

Table 2. Comparison of performance metrics of different classification algorithms

Method	**Precision**	**Recall**	**F1**
WDBP	0.7825	0.8172	0.7995
TSFF	0.8627	0.8731	0.8677
OURs	0.9305	0.9204	0.9254

The accuracy rates of different classification algorithms are compared under different iterations, as indicated in Figure 5. It can be seen that with the increase of the number of training iterations, the classification accuracy of the three algorithms is improved, and although the convergence speed of the OURs model is not as fast as that of the WDBP and TSFF models, the correct rate of the final convergence of the OURs model is significantly higher than that of the other two models, in which the highest accuracy of the OURs

model in the validation set is 0.9831, that of the WDBP model in the validation set is 0.8694, and that of the TSFF model in the validation set is 0.9831. accuracy in the validation set is 0.8694 for the OURs model, and the highest accuracy in the validation set is 0.8762 for the TSFF model.

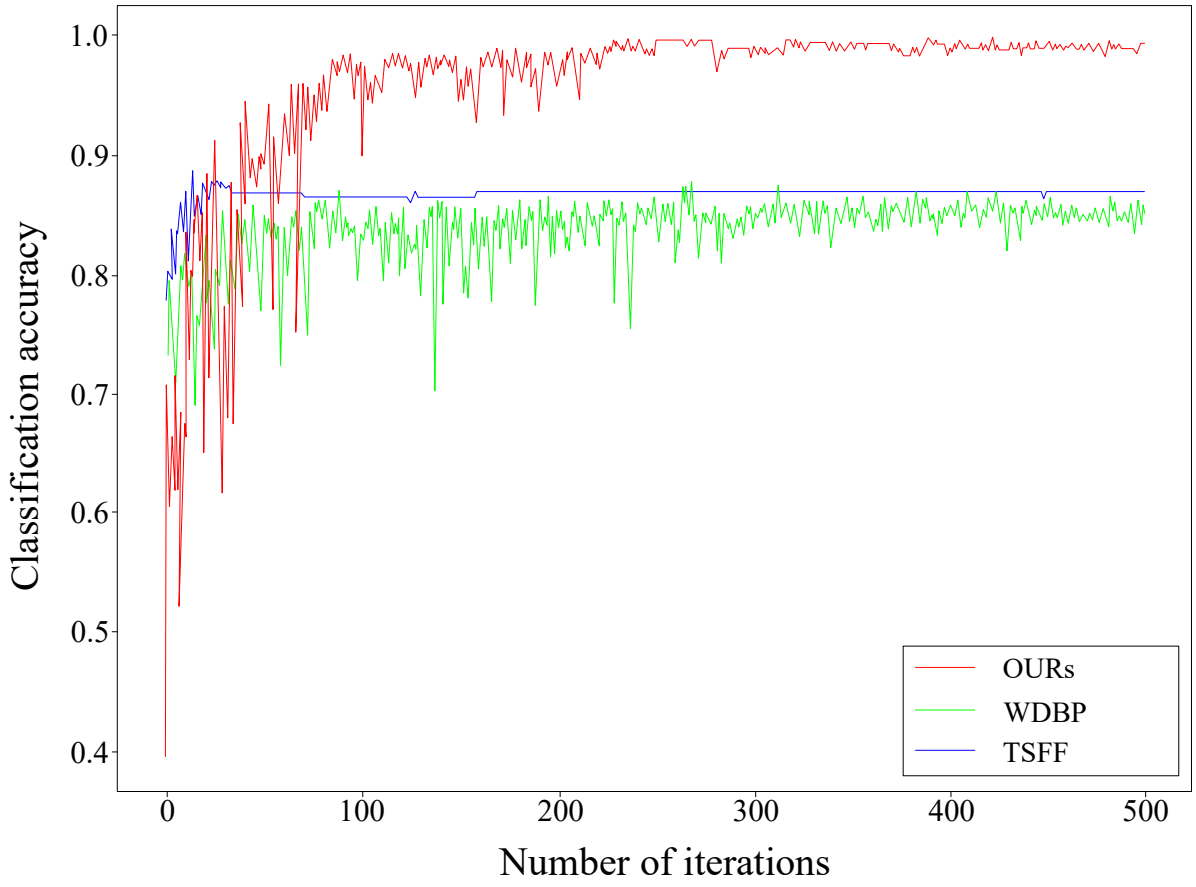

Figure 5. Comparison of accuracy of different classification algorithms

6. **Conclusion.** Intending to the issue of low classification efficiency of current piano soundboard classification algorithms, this article suggests a piano soundboard classification algorithm relied on intelligent neural network and multi-feature fusion algorithm. Firstly, the channel attention mechanism is utilized to share weights to reduce the number of parameters in the CNN model and improve the accuracy of the model. Secondly, the grayscale covariance matrix is adopted to extract the texture features of the soundboard to enhance the texture details, and then the color features of the soundboard image are extracted by adopting the PCA algorithm, and then the texture features and the color features are consistently processed and inputted into the multi-scale convolutional level to obtain the local features, and then the local features are fused with the depth features through the transposed convolutional mapping to obtain the global features. Finally, all the above features are fed into different channels of the ECNN model for feature classification, and the predicted values of each channel are fused as the final prediction results using the weighted average method. The experimental outcome indicates that the suggested algorithm has improved in Accuracy, Precision, Recall and F1 value compared with the comparison methods, indicating that the designed algorithm can be better applied to piano soundboard image classification.

REFERENCES

[1] S.-H. Park, and Y.-H. Park, "Audio-visual tensor fusion network for piano player posture classification," *Applied Sciences*, vol. 10, no. 19, 6857, 2020.

[2] H. Suzuki, "Vibration and sound radiation of a piano soundboard," *The Journal of the Acoustical Society of America*, vol. 80, no. 6, pp. 1573–1582, 1986.

[3] S. Yoshikawa, "Acoustical classification of woods for string instruments," *The Journal of the Acoustical Society of America*, vol. 122, no. 1, pp. 568–573, 2007.

[4] A. Chaigne, B. Cotté, and R. Viggiano, "Dynamical properties of piano soundboards," *The Journal of the Acoustical Society of America*, vol. 133, no. 4, pp. 2456–2466, 2013.

[5] B. Trévisan, K. Ege, and B. Laulagnet, "A modal approach to piano soundboard vibroacoustic behavior," *The Journal of the Acoustical Society of America*, vol. 141, no. 2, pp. 690–709, 2017.

[6] R. M. Haralick, K. Shanmugam, and I. H. Dinstein, "Textural features for image classification," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.

[7] R. F. Walker, P. T. Jackway, and D. Longstaff, "Genetic algorithm optimization of adaptive multiscale GLCM features," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 17, no. 01, pp. 17–39, 2003.

[8] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, 1989.

[9] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, 2002.

[10] I. Y.-H. Gu, H. Andersson, and R. Vicen, "Wood defect classification based on image analysis and support vector machines," *Wood Science and Technology*, vol. 44, pp. 693–704, 2010.

[11] S. Yoshikawa, "Acoustical classification of woods for string instruments," *The Journal of the Acoustical Society of America*, vol. 122, no. 1, pp. 568–573, 2007.

[12] Y. Shi, Y. Li, M. Cai, and X. D. Zhang, "A lung sound category recognition method based on wavelet decomposition and BP neural network," *Int. J. Biological Sciences*, vol. 15, no. 1, 195, 2019.

[13] P. Barmpoutis, K. Dimitropoulos, I. Barboutis, N. Grammalidis, and P. Lefakis, "Wood species recognition through multidimensional texture analysis," *Computers and Electronics in Agriculture*, vol. 144, pp. 241–248, 2018.

[14] S.-Y. Yang, O. Kwon, Y. Park, H. Chung, H. Kim, S.-Y. Park, I.-G. Choi, and H. Yeo, "Application of neural networks for classifying softwood species using near infrared spectroscopy," *Journal of Near Infrared Spectroscopy*, vol. 28, no. 5, pp. 298–307, 2020.

[15] K. Kilic, K. Kiliç, B. B. Sinaice, and U. Özcan, "Wood Species Image Classification Using Two-Dimensional Convolutional Neural Network," *Drvna Industrija*, vol. 74, no. 4, pp. 407–417, 2023.

[16] A. Fabijańska, M. Danek, and J. Barniak, "Wood species automatic identification from wood core images with a residual convolutional neural network," *Computers and Electronics in Agriculture*, vol. 181, 105941, 2021.

[17] Y. Yang, Y. Liu, Z. Liu, and S. Q. Shi, "Prediction of Yueqin acoustic quality based on soundboard vibration performance using support vector machine," *Journal of Wood Science*, vol. 63, pp. 37–44, 2017.

[18] X. Y. Hu, P. L. Wang, and J. Y. Xu, "A wood color classifier based on CAV and SVM," *Applied Mechanics and Materials*, vol. 241, pp. 483–487, 2013.

[19] J.-c. Han, P. Zhao, and C.-k. Wang, "Wood species recognition through FGLAM textural and spectral feature fusion," *Wood Science and Technology*, vol. 55, pp. 535–552, 2021.

[20] Z. Wan, H. Yang, J. Xu, H. Mu, and D. Qi, "BACNN: Multi-scale feature fusion-based bilinear attention convolutional neural network for wood NIR classification," *Journal of Forestry Research*, vol. 35, no. 1, 4, 2024.

[21] T.-Y. Wu, H. Li, S. Kumari, and C.-M. Chen, "A Spectral Convolutional Neural Network Model Based on Adaptive Fick's Law for Hyperspectral Image Classification," *Computers, Materials & Continua*, vol. 79, no. 1, pp. 19–46, 2024.

[22] T.-Y. Wu, A. Shao, and J.-S. Pan, "CTOA: Toward a Chaotic-Based Tumbleweed Optimization Algorithm," *Mathematics*, vol. 11, no. 10, 2339, 2023.

[23] T.-Y. Wu, H. Li, and S.-C. Chu, "CPPE: An Improved Phasmatodea Population Evolution Algorithm with Chaotic Maps," *Mathematics*, vol. 11, no. 9, 1977, 2023.

[24] F. R. De Siqueira, W. R. Schwartz, and H. Pedrini, "Multi-scale gray level co-occurrence matrices for texture description," *Neurocomputing*, vol. 120, pp. 336–345, 2013.

[25] F. Zhang, T.-Y. Wu, Y. Wang, R. Xiong, G. Ding, P. Mei, and L. Liu, "Application of Quantum Genetic Optimization of LVQ Neural Network in Smart City Traffic Network Prediction," *IEEE Access*, vol. 8, pp. 104555–104564, 2020.

[26] M. C. Cooper, "An optimal k-consistency algorithm," *Artificial Intelligence*, vol. 41, no. 1, pp. 89–95, 1989.

[27] F. Zhang, T.-Y. Wu, and G. Zheng, "Video salient region detection model based on wavelet transform and feature comparison," *EURASIP Journal on Image and Video Processing*, vol. 2019, pp. 1–10, 2019.

[28] F. Zhang, Y. Wang, and C. Wu, "An automatic generation method of cross-modal fuzzy creativity," *Journal of Intelligent & Fuzzy Systems*, vol. 38, no. 5, pp. 5685–5696, 2020.

[29] A. Berger, and S. Guda, "Threshold optimization for F measure of macro-averaged precision and recall," *Pattern Recognition*, vol. 102, 107250, 2020.