Classification of Optical Music Symbols based on Combined Neural Network

Cuihong Wen*, Ana Rebelo†, Jing Zhang‡, and Jaime Cardoso§
*College of electrical and information engineering, Hunan University, Changsha, 410082
Email: cuihongwen2010@163.com
†INESC Porto, Universidade do Porto, Porto, Portugal
Email: arebelo@inescporto.pt
‡College of electrical and information engineering, Hunan University, Changsha, 410082
Email: zhangj@hnu.edu.cn
§INESC Porto, Universidade do Porto, Porto, Portugal
Email:jaime.cardoso@inescporto.pt

***Abstract*—In this paper, a new method for music symbol classification named Combined Neural Network (CNN) is proposed. Tests are conducted on more than 9000 music symbols from both real and scanned music sheets, which show that the proposed technique offers superior classification capability. At the same time, the performance of the new network is compared with the single Neural Network (NN) classifier using the same music scores.**

I. INTRODUCTION

Optical Music Recognition (OMR), which is an important tool to recognize a scanned page of music symbols automatically, has received increasing attentions in the past few decades [1]. The OMR is important because most of produced musical works in the past are still available only as original manuscripts or photocopies, while the preservation of these works requires their digitalization and transformation into a machine readable format. An OMR program is able to recognize the musical content and make the semantic analysis of each musical symbol of a musical work. In addition, it makes the searching, retrieving and analyzing of the music sheets easier. Thus, it is regarded as one of the most promising tools to preserve the music scores.

Most of the recent work on OMR focused on staff detection and removal[2][3], and the music symbol segmentation[4][5]. Besides, there are also a great number of works focused on the music symbols classification[1][6] using certain methods, such as the k-nearest neighbor (KNN) and support vector machine (SVM). Furthermore, in[7], the segmentation and classification were performed simultaneously using Hidden Markov Models (HMMs), which showed that the operation of symbol classification can be linked with the segmentation of the objects from the music symbols.

However, although each of these approaches has been shown to be effective in specific environments, the results of the classification of music scores are still far from ideal. In this paper, a method for music symbol classification in handwritten and printed scores will be present. The Combined Neural Network (CNN) which is believed has the potential to achieve a better recognition accuracy, will be used as the classifier for classifying music symbols.

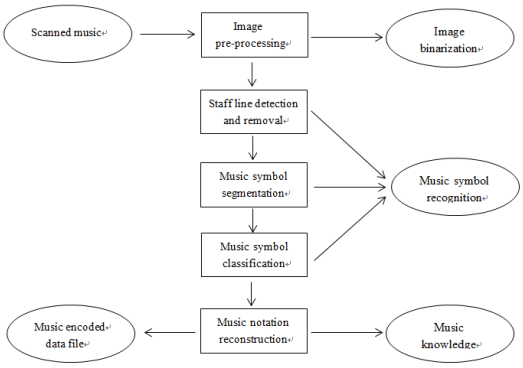

Fig. 1: How OMR can be simplified into five smaller tasks

The remainder of this paper is structured as follows. In section 2 we review a general framework that decomposes the problem of OMR into key stages. This part includes image pre-processing, staff line detection and removal, music symbol segmentation, music symbol classification and music notation reconstruction. In section 3 we describe the proposed algorithm for classifying the music symbols, highlight architecture of the combined neural network, and give the details of the database we used. In the section 4 we present the results and compare the results with the other network. Finally, Section 5 concludes the paper and proposes future work.

II. SYSTEM ARCHITECTURE FOR OMR

In principle OMR is an extension of Optical Character Recognition (OCR). However, the problems to be faced are more complex due to the connection of normally separated primitives or broken symbols. Thus, as in Fig. 1 the OMR is simplified through decomposition. Generally, the process is divided into five steps. It includes image pre-processing, staff line detection and removal, music symbol segmentation, music symbol classification, and music notation reconstruction. Such tasks are challenging and require the development and integration of techniques from diverse areas, including computer vision, artificial intelligence, machine learning, and music theory.

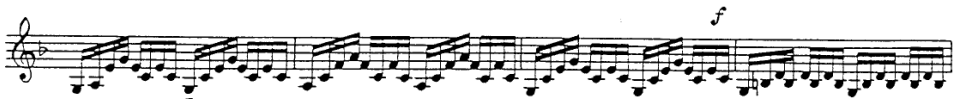

Fig. 2: Before staff line removal

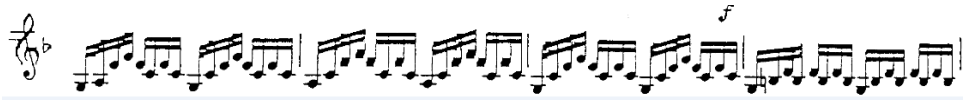

Fig. 3: After staff line removal

A. Image pre-processing

It consists the application of several techniques (e.g. binarization, noise removal, blurring, deskewing, etc.) to make the recognition process more robust and efficient.

B. Staff line detection and removal

Staff line detection and removal are one of the fundamental stages on the OMR process, with subsequent processes relying heavily on their performance. The goal of staff line removal is to remove the lines as much as possible while leaving the symbols on the lines intact. This task dictates the possibility of success for the recognition of the music score. Fig.3 is an example of staff line removal for Fig.2.

C. Music symbol segmentation

This segmentation process consists in localizing and isolating the symbols in order to identify them. In [6], the symbols are split into several different types:notes,beams,clefs, accents,etc. The segmentation of these types of symbols was based on a hierarchical decomposition of a music image. A music sheet is first analyzed and split by staffs, as yielded by the staff lines removal step. Subsequently, the series connected components were identified. To extract only the symbols with appropriate size, a series selection of the connected components detected in the previous step was carried out. The thresholds used for the height and width of the symbols were experimentally chosen. These values take into account the features of the music symbols. As a bounding box of a connected component can contain multiple connected components, care was taken in order to avoid duplicate detections or miss to detect any connected component. In the end, we are ready to find and series extract all the music symbols.

D. Music symbol classification

Once the symbols have been segmented, the challenge is to classify them. At this step,several sets of symbols are extracted from different musical scores to train the classifiers. Then the symbols are grouped according to their shape and a certain level of music recognition has been accomplished. Table I is the classes list of handwritten and printed music symbols.

Techniques from the area of Document Image Analysis that have been successfully adapted and applied to OMR to solve the music symbols classification stage.There are some straightforward classifiers for musical symbol classification

TABLE I: Full set of the handwritten and printed music symbols considered.

Accent	BassClef	Beam	Flat	natural
Note	NoteFlag	NoteOpen	RestI	RestII
Sharp	TimeN	TrebleClef	TimeL	AltoClef
Relation	Breve	Semibreve	Dots	Barlines

including Hidden Markov models,K-nearest neighbor,Neural networks,Support vector machines,etc.As described in [6]they conducted a comparative study of classification methods for musical primitives and examined four classification methods.

Hidden Markov Model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (hidden) states. One of the reasons for the use of HMM lies in the capability to perform segmentation and recognition at the same time[7].

The k-nearest neighbor algorithm is amongst the simplest of all machine learning algorithms. This algorithm belongs to a set of techniques called Instance-based Learning. It starts by extending the local region around a data point until the kth nearest neighbor is found. An object is classified by a majority vote scheme, with the object being assigned to the class most common amongst its k-nearest neighbors. The training lies only in the estimation of the best k.

Support vector machine (SVM), follows the main idea of constructing a hyperplane as the decision surface in such a way that the margin of separation between positive and negative examples is maximized. It is supervised learning model with associated learning algorithms which can be used for classification.

Artificial neural network, or neural network for short, was originally inspired on the central nervous systems and on the neurons, which constitute one of its most significant information processing elements. With time, it has evolved quite independently from the biological roots, giving rise to more practical implementations, based on statistics and signal

processing. In our days, several applications have been found based on the principles and algorithms of neural networks. The focus of this paper is the classification of the music symbols with a designed combined neural network(CNN) and we give the architecture of the CNN in section III.

E. Music notation reconstruction

Detected symbols are interpreted and assigned a musical meaning. The relationship between symbols is determined and the information is stored in a form that programs such as sequencers or music editors can use the symbols primitives are merged to form musical symbols. A format of musical description is created with the information previously produced. Usually, in this step, graphical and syntactic rules are used to introduce context information to validate and solve ambiguities from the previous module of music symbol recognition[8].

III. PROPOSED ARCHITECTURE OF THE COMBINED NEURAL NETWORK(CNN)

A theory of classifier combination of Neural Network was discussed in[9]. Our CNN is based on the thoery of [9]. Our CNN is based on the theory of[9]. The main idea behind is to combine decisions of individual classifiers to obtain a better classifier. To make this task more clearly defined and subsequent discussions easier, the architecture of the CNN is described in Fig.4.

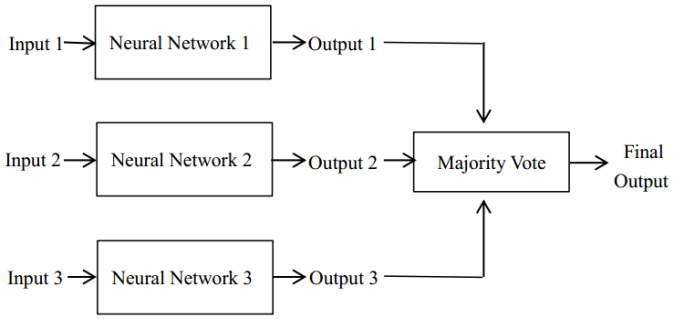

Fig. 4: The structure of the CNN

The three Neural Networks are the same network named Multi-layer Perception(MLP),which will be introduced in the following subsection. And the other focus of CNN is the study how the amount of information presented in output vectors affect combined performance. This can be easily achieved by applying different majority vote functions.

A. The inputs

First,each music symbol image was converted to binary image by thresholding.Then we resize the images.For input1 and input 2,the images were resized to 20×20 pixels and then converted to a vector of 400 binary values.At the same time,we give the images of the input 3 a different size,in which case the images were resized to 30×40 pixels and then converted to a vector of 1200 binary values.

B. Database

A data set of both real handwritten scores and synthetic scores was adopted to perform the CNN. The real scores consist on a set of 65 handwritten scores from 6 different composers. As mentioned,the input images were previously binarized with the Otsu threshold algorithm. In the synthetic data set, a number of distortions were applied. This set consists on the fraction of the dataset, available from [10], written on the standard notation. The deformations applied to these printed scores were curvature, rotation, Kanungo and white speckles,see[10] for more details. In total, 380 distorted images were generated from 19 original scores. The relevant classes for handwritten/printed music symbols used in the training phase of the classification models are presented in Table I. The symbols are grouped according to their shape. The rests symbols were divided into two groups RestI and RestII. In total the classifiers were evaluated on a database containing 7128 examples divided into 20 classes.

C. Multi-layer Perception (MLP)

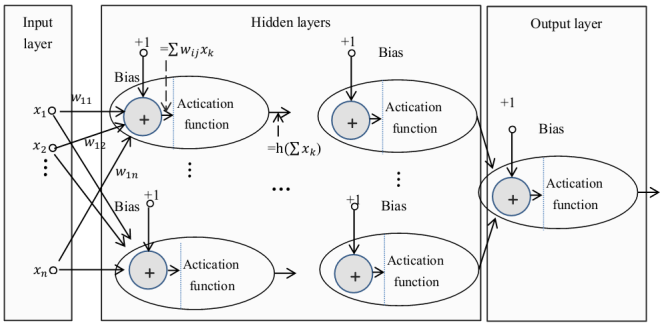

Fig. 5: The structure of the MLP

Multi-layer Perception (MLP), one type of a feed-forward neural network that have been used in pattern recognition problems as early as 1957 when Rosenblatt introduced the perception[11]. The network is composed of layers consisting of various number of units. Units in adjacent layers are connected through links whose associated weights determine the contribution of units on one end to the overall activation of units on the other end.

There are generally three types of layers. Units in the input layer bear much resemblance to the sensory units in a classical perception. Each of them is connected to a component in the input vector. The output layer represents different classes of patterns. Arbitrarily many hidden layers may be used depending on the desired complexity. Each unit in the hidden layer is connected to every unit in the layer immediately above and below. A diagram of a basic MLP network is shown in Fig.5.

The training of the networks was carried out under Matlab 7.8. As mentioned above, the inputs are the vectors of binary values converted from the music symbol images. We use a network with K outputs, and each one represents the corresponding class of each image. Further more,we saved the probabilities for each image being classified to each class.

D. Experimental testing

For evaluation of the pattern recognition processes, the available dataset was randomly split into three sub-sets: training, validation and test sets, with25%, 25% and 50% of the data, respectively. This division was repeated 4 times in order to obtain more stable results for accuracy by averaging and also to assess the variability of this measure. No special constraint was imposed on the distribution of the categories of symbols over the training, validation and test sets. we only guaranteed that at least one example of each category was present in the training set.

In this work the CNN classifiers were tested using test sets randomly generated above. And we could get the final results by majority vote of the three outputs of CNN.

E. Majority vote

As showed in Fig.4, there are three different outputs for three Neural Networks. The combined performance depends on the choosing of the method for majority vote. We have applied two different majority vote methods.

The first method we label it as CNNMV1. In this case,we save the three outputs of the NN1,NN2,and NN3 together in a matrix named M. Then make a decision using the following algorithm.

Algorithm 1 Algorithm of CNNMV1

```
if length(unique(M(:,j)))==1, which means the values of the
jth column of M are same then
   we choose this value as the output,
else
   if length(unique(M(:,j)))==2, which means two of the
   values of the jth column of M are same then
      we choose the majority one as the output,
   end if
else
   if length(unique(M(:,j)))==3, then
      we check the probability matrix and choose the biggest
      probability and use the index to make a decision,
      [val,idx]=max(PROB(:,j));
      class(1,j)=M(idx,j).
   end if
end if
```

After that, we have a class vector in which each value stands for the class of related image. Then we repeat four times with different test sets randomly generated and save the four class vectors as a matrix CLASS. At last, we calculate the errors by comparing the values of each row of the CLASS with the target class we saved at the beginning,and choose the raw of the CLASS which generates minimal error as the final output. The accuracy of the CNNMV1 is showed in Table II.

The second one we label it as CNNMV2. CNNMV2 is much easier than the CNNMV1. CNNMV2 is much easier than the CNNMV1. Because we had three outputs of the NNs, and we repeated four times of CNN, we finally had twelve classification results. The main idea of CNNMV2 is to save all the twelve results vectors together in a matrix and choose the most frequency value for the final output.Then calculate the error and accuracy.

TABLE II: The Results of NN and CNNMV1 and CNNMV2

Classes	Accuracy of NN	Accuracy of CNNMV1	Accuracy of CNNMV2
Accent	82%	91%	97%
BassClef	92%	98%	98%
Beam	91%	96%	100%
Flat	82%	91%	100%
Natural	89%	96%	99%
Sharp	95%	100%	100%
TimeN	96%	100%	100%
TrebleClef	94%	99%	100%
TimeL	96%	98%	98%
AltoClef	87%	96%	100%
Note	78%	78%	93%
NoteFlag	86%	94%	95%
NoteOpen	81%	95%	99%
RestI	88%	96%	100%
RestII	87%	98%	100%
Relation	76%	84%	100%
Breve	89%	97%	97%
Semibreve	94%	100%	100%
Dots	82%	97%	99%
Barlines	90%	99%	100%
Average Accuracy	88.04 %	95.67%	98.82%

IV. RESULTS USING PROPOSED CNN

A. Results using proposed neural network

Tables II presents the results obtained applying NN and CNN classifiers in the OMR database,using both of the majorty vote methods proposed in the this paper. For a test set of 2253 music symbol images,an accuracy of over 98 percent was reached.

The first assessment is that within CNNMV1 methodology, an overall improvement was observed. Moreover, CNNMV2 achieved the best results where the average accuracy reached 98.82%. Our approach of CNNMV2 seems to work extremely well for the purpose of music symbol classification. Due to different applications, the training stages, and the testing sets of data, comparison between the performance of our proposed network and those of the others mentioned is difficult. However, based purely on the accuracy results, our network seems to out perform any of these networks.

B. Compare results with the other network

The network that we have been using for our existing prototype is the only one Neural Network(NN). When we compared NN with the CNN presented in this paper, a marked improvement in recognition accuracy was observed, using the same system setup and dataset.

V. CONCLUSIONS AND FUTURE WORK

In this article a CNN classifier was successfully applied to recognize music symbols. Significant classification improvements were obtained. Further investigations could include combining other classification models using similar method of this paper and finding more majority vote algorithms to make the

classification performance better. This line of research involves a study and a profound understanding of the latest techniques of pattern recognition, machine learning. The merger of rules and techniques from different areas may help improve the existing algorithms.

ACKNOWLEDGMENT

This work is financed by Fund of Doctoral Program of the Ministry of Education (Approval No.20110161110035) and China National Natural Science Foundation (Approval No. 61174140,61203016 and 61174050).

REFERENCES

[1] Ana Rebelo, I. Fujinaga, F. Paszkiewicz, C. Guedes, A. Marcal, and J. Cardoso, "Optical music recognition: State of the art and open issues for handwritten music scores" International Journal of Multimedia Information Retrieval, Volume 1, no. 3, pp.173-190, October 2012.

[2] I Fujinaga.Staff Detection and Removal. In S. George (editor), Visual Perception of Music Notation, 1-39,2004.

[3] J. S. Cardoso, A. Capela, A. Rebelo, C. Guedes, and J. P. da Costa, Staff detection with stable paths,IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 6, pp. 1134C1139, 2009.

[4] P. Bellini, I. Bruno, and P. Nesi, "Optical music sheet segmentation," Proceedings of the 1st International Conference on Web Delivering of Music, pp. 183-190, Florence, Italy, IEEE Computer Society Press, November 2001.

[5] F. Rossant and I. Bloch, "Robust and adaptive OMR system including fuzzy modeling, fusion of musical rules, and possible error detection," EURASIP Journal on Advances in Signal Processing, no. 081541, pp.160-160, 2007.

[6] Ana Rebelo, Artur Capela, and Jaime S. Cardoso, "Optical Recognition of Music Symbols: a comparative study," International Journal of Document Analysis and Recognition, Vol. 13, no. 1, pp.19-31, March 2010.

[7] L. Pugin, "Optical music recognition of early typographic prints using Hidden Markov Models," International Society for Music Information Retrieval (ISMIR), pp. 53-56, 2006.

[8] K. C. Ng, R. D. Boyle, "Recognition and reconstruction of primitives in music scores," Vol 14, no. 1, pp. 39-46, February 1996.

[9] Dar-Shyang Lee, "A theory of classifier combination: the neural network approach," Third International Conference on Document Analysis and Recognition, ICDAR 1995, August 14 - 15, 1995.

[10] C. Dalitz, M. Droettboom, B. Czerwinski, and I. Fujigana, "A comparative study of staff removal algorithms," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, pp. 753-766, May 2008.

[11] F. Rosenblatt, The perceptron: A perceiving and recognizing automaton. Cornell Aeronaut. Lab Report, 85-4601-1, 1957.