CREATES - Convolutional Neural Network applied to Identification of Musical Elements in OMR

Jenaro Augusto Barbosa*, Edimilson Batista dos Santos

[1]Universidade Federal de São João del-Rei – UFSJ
Departamento de Ciência da Computação – DCOMP
Campus Tancredo Neves (CTAN) – São João del-Rei, MG

jenaro@ufsj.edu.br, edimilson.santos@ufsj.edu.br

Abstract. *Optical music recognition (OMR) is an important tool to recognize a scanned page of music sheet automatically, which has been applied to preserving music scores. In this paper, we present a comparative study among Convolutional Neural Network (CNN) and state-of-the-art classifiers to classify musical symbols. The initial results show that the CNN is promising in this task according to provided information.*

1. Introduction

A significant amount of musical works produced in the past are still available only as original manuscripts or as photocopies on date. In several historical cities, it is possible to find important music collections dated from the beginning of 18th century which are in this state. Thus, some works, such as Copista system [1], have proposed to apply computer science techniques to musical scores recognition through a process named Optical Music Recognition (OMR).

Optical Music Recognition (OMR) is a Computer Science field applied to music that deals with problems like recognition of handwritten scores. The applications in OMR are similar to Optical Character Recognition (OCR) tools. However, it is not a straightforward extension from the OCR since the problems to be faced are substantially different.

The OMR is needed for the preservation of musical works which requires digitalization and should be transformed into a machine readable format. An OMR program should thus be able to recognize the musical content and make semantic analysis of each musical symbol of a musical work. Generally, such a task is challenging because it requires the integration of techniques from some quite different areas, i.e., computer vision, artificial intelligence, machine learning, and music theory. In spite of existing applications that converts handwriting scores into editable scores, most of these applications a) do not work with manuscript scores[2],b) are very expensive and c) are not open source. All these reasons encourage to build a brand new tool on the OMR. The development of an OMR system can be divided into some distinct parts, e.g.: the image acquisition, image preprocessing and digital image recovery, the recognition of musical symbols with computer vision and machine learning, the music notation reconstruction and the symbolic music output. In this paper, we focus on the step of recognition of musical symbols through machine learning techniques. Specifically, we propose to apply classification algorithms to identify such symbols, with emphasis on Convolutional Neural Networks, since it can obtain good results without many image processing steps.

The **CREATES** - **C**onvolutional neu**R**al n**E**twork **A**pplied **T**o id**E**ntification mu**S**ical - solution can be divided into some stages. The separation occurs in four parts: (A) — raw state, (B) — preprocessing, (C) — CNN, (D) — Output, being a base solution for a future framework to be later implemented.

*Supported by 004/2019/PROPE/UFSJ.

2. Related Work

Many musical works produced in the past are still currently available only as original manuscripts or as photocopies. The OMR is needed for the preservation of these works which requires digitalization and should be transformed into a machine readable format. Despite the many research activities on optical music recognition (OMR), the results for handwritten musical scores are far from ideal.

A lot of work on the OMR include staff lines detection and removal [3, 4, 5, 6], music symbol segmentation [7, 8], a tool proposal to convert handwriting scores into a digital music representation [1].

For many works, the machine learning area offers several methods to classify the music symbols. In [9], the authors propose a new combined neural network classifier, which has the potential to achieve a better recognition accuracy. In [10], a comparative study of several recognition algorithms of music symbols is presented. Although all the approaches have been shown to be effective in specific environments, they all suffer from some limitations. Thus, in this paper, we propose to apply Convolutional Neural Networks (CNNs) (see 3.1) in the task of classifying musical symbols for OMR systems. CNNs have obtained good results in image analysis and it can do without many image processing steps.

3. CREATES - CNN proposed to Recognize Musical Elements

Fig. 1 briefly presents the stages of the proposed solution for the problem of recognition of musical elements. The first two stages in Fig. 1, (A) and (B), are applied to the other classifiers and to the CREATES, being a part of the solution.

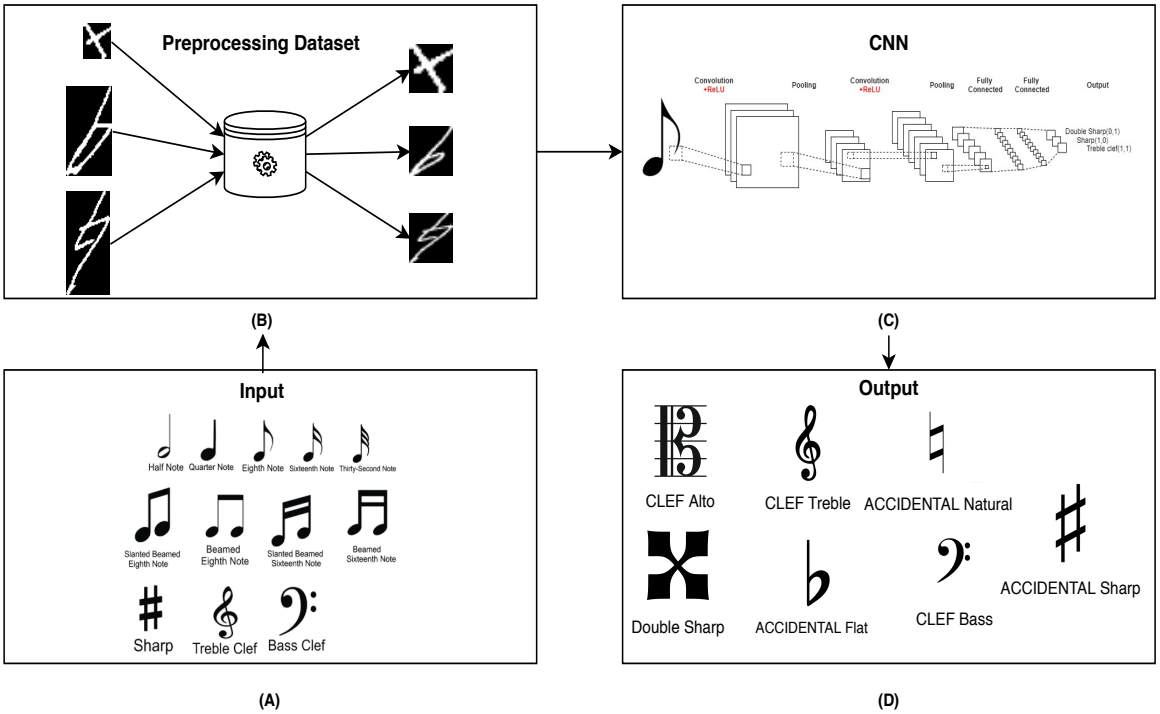

Figure 1: CREATES - Flowchart of the proposed solution.

The stage (A) - raw state - is the first stage with images in the initial state from the dataset [11]), which has 4094 instances of 7 types of musical symbols. These symbols can be seen in Figure 2. The images in the set have different dimensions and the set is not separated into a test and training set.

Figure 2: Each of the seven types of musical symbols used in the initial experiments.

The second stage is (B), where pre-processing in the images was performed to define a standard size (28x28). Here, the OpenCV library is used for this and also an interpolation process to resize. Next, a random separation process is carried out to select and generate training and test datasets. Thus, the original dataset was divided in 80% for training and 20% for testing. These pre-processed datasets were used in the experiments with the classifiers. Lastly, a numerical value was assigned to each musical symbol to label the classes (destination). Table 1 shows the number of instances for each musical symbol in the training and test datasets, after pre-processing .

The proposed solution, CREATES, comprises the use of the Convolutional Neural Network (CNN) for the classification of musical notes.

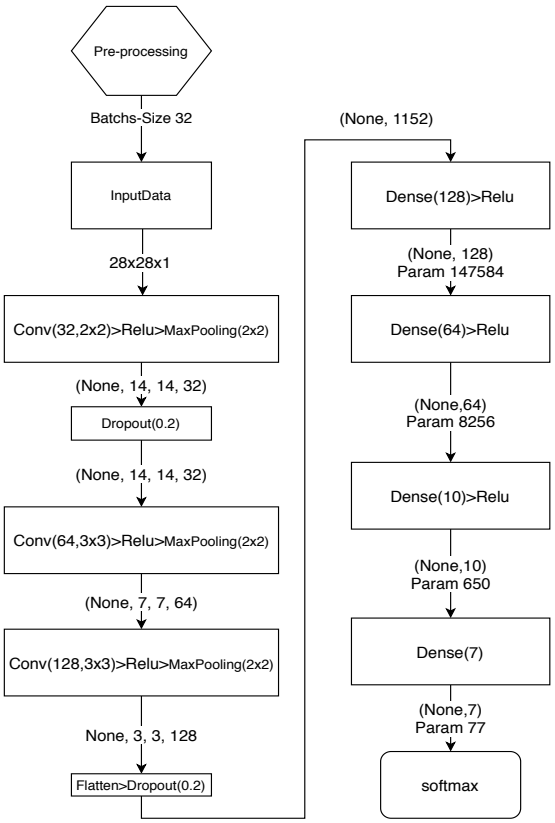

Figure 3: Proposed CNN architecture.

Table 1: Datasets to run the classification algorithms

Symbols	Instances[1]	Train	Test	Class
Alto	759	607	152	0
Double Sharp	497	397	100	1
Treble	820	656	164	2
Flat	517	413	104	3
Natural	471	376	95	4
Bass	549	439	110	5
Sharp	481	384	97	6
Total	**4094**	**3272**	**822**	

[1] Instance: number of instances for each symbol.

3.1. Proposed CNN architecture

In the third stage of the solution presented in Fig.1 (C), a CNN architecture was developed from the open source library known as TensorFlow to classify a set of musical symbols. The CNN's training and test have been performed on Google Colaboratory. For the training, the images have been divided in batch with size of 32 an use the Keras .

The CNN architecture is presented in Fig. 3. It can be divided in two parts: i) features extraction and ii) set of dense layers. In the features extraction step, the CNN architecture has 3 layers of convolution, 3 layers of polling, and 1 regularization layer. The parameters, such as number of filters and kernel size, have been defined empirically. In all of these layers, the ReLu activation function is used, since it is popular in works of the area [7]. The regularization layer (dropout(0.2)) tries to avoid the overfitting, eliminating 20% of the output units from the previous polling layer.

In the second part of the CNN architecture, we have the dense layers in addition to 1 transformation layer and 1 regularization layer. The first layer in the second part is of the Flatten type, which can be understood as a separation between the two parts of the network. Usually, this layer operates a transformation on the image matrix, changing it to an array. The regularization layer also applies a 20% discard rate on the value of the previous layer's output. All the dense layers use the ReLu activation function, except the last that uses the Softmax activation function. This CNN model has been compiled using the Adam optmization function, which is a stochastic gradient descent method, based on the adaptive estimation of first and second order moments.

4. Initial Experiments and Analysis of Results

In the initial experiments, we used CRATES in addition to the following state-of-the-art classifiers: SVM, KNN, Neural Network MLP, Gaussian Naive Bayes, Multinomial Naive Bayes, Complement Naive Bayes and Bernoulli Naive Bayes. For these classifiers, we use the Scikit-learn library and the same platform Google Colab, using the training and test datasets presented in the first and second stages of the solution Figure 1.

The parameters of some algorithms were empirically defined using the randomized searchCV function available in the Scikit-learn library. Thus, the KNN has obtained better results with the number of neighbors K = 2. A multi-layer neural network was used, containing 5 hidden layers and 100 neurons in each. For the other algorithms applied in the experiments, the parameters remained with the default values defined in the library.

Table 2 shows the results of the classifiers when trained with grayscale images, using comparison metrics such as accuracy, precision, recall and F1-Score. CNN presented the best results (**92 %** of accuracy, **93 %** of precision, **92 %** of recall and **92 %** of F1-Score), presenting greater relation to the others classifiers, but with a longer time than the other classifiers. In this experiment, the SVM algorithm showed a second longer time compared to others classifiers, but it obtained low accuracy (**20%**), low recall (**14%**) and low F1-Score (**5%**) when compared to other state-of-the-art classifiers. We believe that the SVM algorithm was not able to learn the using image dataset in grayscale.

Table 2: Results (in %) obtained by classification algorithms using images

Classifiers	Acc	Pre	Recall	F1	Time
SVM	20	89	14	5	28.09s
KNN	64	65	65	64	4.86s
MLP	66	66	65	66	6.85s
Gaussian NB	60	60	62	60	0.068s
Multinomial NB	63	64	63	63	0.083s
Complement NB	55	59	59	58	0.078s
Bernoulli NB	63	64	65	64	0.097s
CNN	**92**	**93**	**92**	**92**	**120s**

As the accuracy may not be so reliable, since the classes do not have the same number of instances, Fig. 4 presents the confusion matrix obtained by proposed CNN. It is possible to note that the CNN maintains a good hit rate for each class, only with more errors in the flat class.

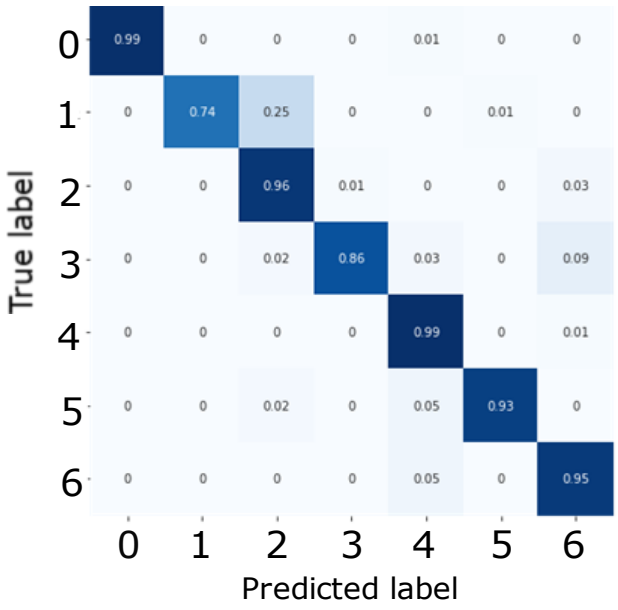

Figure 4: Confusion matrix obtained by proposed CNN architecture. Label Confusion Matrix: 0-Double Sharp, 1-Flat, 2-Natural, 3-Sharp, 4-Alto, 5-Bass, 6-Treble.

5. Conclusions

In this paper, it is proposed the investigation of algorithms and classification methods that can help in the recognition of musical elements for OMR applications. The initial proposal is to apply Convolutional Neural Networks, which is able to obtain good results in image analysis without performing much preprocessing.

The initial experiments provided results for the comparison of different classifiers. The proposed CNN proved to be adequate for the task of recognizing musical symbols.

According to the literature, it is noticed that the area of OMR (Character Recognition) manuscripts is still new and full of obstacles. The results obtained in this paper indicate that there is much to be done in this area. Thus, it is intended to continue the development of this project, taking advantage of the experience and the results acquired and applying to the development of a system to convert handwriting scores into a digital music representation.

References

[1] Avner Maximiliano de Paulo, Flávio Luiz Schiavoni, Marcos Antônio de Matos Laia, and Daniel Luiz Alves Madeira. Copista-sistema de omr para a recuperaç ao de acervo histórico musical. *XV SBCM-Computer Music: Beyond the frontiers of signal processing and computational models*, 2015.

[2] David Bainbridge and Tim Bell. The challenge of optical music recognition. *Computers and the Humanities*, 35(2):95–121, 2001.

[3] Ana Rebelo and Jaime S Cardoso. Staff line detection and removal in the grayscale domain. In *2013 12th International Conference on Document Analysis and Recognition*, pages 57–61. IEEE, 2013.

[4] Christoph Dalitz, Michael Droettboom, Bastian Pranzas, and Ichiro Fujinaga. A comparative study of staff removal algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(5):753–766, 2008.

[5] Ichiro Fujinaga. Staff detection and removal. In *Visual Perception of Music Notation: On-Line and Off Line Recognition*, pages 1–39. IGI Global, 2004.

[6] Jaime dos Santos Cardoso, Artur Capela, Ana Rebelo, Carlos Guedes, and Joaquim Pinto da Costa. Staff detection with stable paths. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(6):1134–1139, 2009.

[7] Florence Rossant and Isabelle Bloch. Robust and adaptive omr system including fuzzy modeling, fusion of musical rules, and possible error detection. *EURASIP Journal on Advances in Signal Processing*, 2007(1):081541, 2006.

[8] Alicia Fornés, Josep Lladós, and Gemma Sánchez. Primitive segmentation in old handwritten music scores. In *International Workshop on Graphics Recognition*, pages 279–290. Springer, 2005.

[9] Cuihong Wen, Ana Rebelo, Jing Zhang, and Jaime Cardoso. A new optical music recognition system based on combined neural network. *Pattern Recognition Letters*, 58:1–7, 2015.

[10] Ana Rebelo, G Capela, and Jaime S Cardoso. Optical recognition of music symbols. *International Journal on Document Analysis and Recognition (IJDAR)*, 13(1):19–31, 2010.

[11] Alicia Fornés, Josep Lladós, and Gemma Sánchez. Old handwritten musical symbol classification by a dynamic time warping based method. In *International Workshop on Graphics Recognition*, pages 51–60. Springer, 2007.