INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.165

Automatic Music Composition Using AI

Nikita M Survase, Akanksha S Popade, Rohan B. Ghorpade, Dhiraj P. Bhosale,

Prof. Laxmikant Malphedwar

Department Computer Engineering, Dr. D. Y. Patil college of Engineering and Innovation, Varale, Talegaon,

Pune, India.

ABSTRACT: Algorithmic music composition is concerned with generation of music through algorithms. Using natural language text as the basis for generation opens up interesting ideas for research. We have identified three strategies for mapping text to music: using all the letters, using only vowels and using vowels in conjunction with the word's part of speech (POS). The third approach is best achieved with a full parse tree since it helps in identifying the most probable POS tag. Using text input and applying these strategies, we derive a sequence of notes (pitches) as well as the rhythm (note durations).

KEYWORDS*:* duration, music, conjunction

I. INTRODUCTION

The recognition of music scores [1, 2, 3] has attracted the interest of the research community for decades. Since the first works in the 60s [4] and 70s [5], the recognition or music scores has significatively improved. In the case of printed music scores, one could say that the state of the art has reached a quite mature state. As a matter of fact, many commercial OMR systems show very good performance, such as PhotoScore [6] or SharpEye [7]. Nowadays, a lot of handwritten music scores exist only in the form of original or photocopied manuscripts, without.being published in a digital format. Digitization is an essential tool to preserve this cultural heritage. For this, Optical Music Recognition (OMR) algorithms are required. An OMR system has a previous image preprocessing stage (application of techniques such as enhancement, binarization, noise removal, among others). A typical OMR framework has three principal modules [8]: (1) recognition of musical symbols. The staff lines have to be removed and the primitives detected. (2) Reconstruction of the musical information. Here its common use graphical and syntactic rules to overcome possible classification errors and (3) construction of a musical notation model for its representation which can be a MIDI, digital music score, etc. Concerning handwritten scores, although it is remarkable the work in Early musical notation [9, 10], the recognition of handwritten Western Musical Notation still remains a challenge. The main two reasons are the following. First, the high variability in the handwriting style increases the difficulties in the recognition of music symbols. Second,the music notation rules for creating compound music notes (groups of music notes) allow a high variability in appearance that require special attention. In order to cope with the handwriting style variability when recognizing individual music symbols (e.g. clefs, accidentals, isolated notes), the community has used specific symbol recognition methods [11, 12] and learning-based techniques such as SVMs, HMMs or ANNs [13]. As stated in [14], in the case of the recognition of compound music notes, one must deal not only with the compositional music rules, but also with the ambiguities in the detection and classification of graphical primitives (e.g. headnotes, beams, stems, flags, etc.). It is true that temporal information is undoubtedly helpful in on-line music recognition, as it has been shown in [15, 16]. Nowadays, a musician can find several applications for mobile devices, such as StaffPad [17], MyScript Music [18] or NotateMe [19]. Concerning the off-line recognition of handwritten groups of music scores, much more research is still needed. PhotoScore seems to be the only software able to recognize off-line handwritten music scores, and its performance when recognizing groups of notes is still far from satisfactory. One of the main problems is probably the lack of sufficient training data for learning the high variability in the creation of groups of notes.

II. LITERATURE REVIEW

Visual object detection is a very active field of research with remarkable results on detecting objects in natural images with a variety of active competitions. Many competing approaches have been proposed in the last few years such as Faster R-CNN [4], R-FCN [5] and Single shot detectors [6], [7]. While some optimize for accuracy, others strive for high performance [8]. However, all of them share the fact, that heavily make use of deep convolutional neural

networks. The traditional pipeline of segmenting and classifying symbols has been shown to work well on simple typeset music scores with a known music font [9]. But when considering low-quality images, complex scores or even handwritten ones, these systems tend to fail, mainly because errors propagate from one step to subsequent steps [10], e.g. a segmentation error could cause incorrectly detected objects. Initial attempts to overcome this limitation by directly detecting music objects with CNNs were made by [11], who suggest an adaptation of Faster RCNN with a custom region proposal mechanism based on the morphological skeleton to accurately detect noteheads and [12], who are able to detect accidentals in dense piano scores with high accuracy, given previously detected noteheads, that are being used as input-feature to the network. However, both of them are limited to experimentations on a tiny subset of the full vocabulary used in modern music notation. Although both approaches can be extended to other classes, it remains an open question, whether a general purpose detector that can learn a large vocabulary is superior to multiple class-specific detectors. A very interesting alternative to the traditional OMR pipeline is the attempt of solving OMR in a holistic fashion. The first notable attempt at doing so was by Pugin [13], who used Hidden Markov Models to read typographic prints of early music. More recently, the combination of using CNNs jointly with Recurrent Neural Networks to build an end-to-end trainable OMR system [14] was adapted and extended by [15] and [16]. Both train very similar models on a very large set of monophonic music scores containing a single staff per image. Although the reported results on the given datasets are very good, these systems currently exhibit the following limitations: • They operate only on very primitive, printed, monophonic scores. Extending their pipeline to more complex music scores with multiple voices requires a different formulation of the output data to at least include onset and offset of each note and not only the pitch and duration. • By using pooling operations during the feature extraction, the network gains location invariance that conflicts with the interest of precise location information, which is needed to correctly infer the pitch of a note. • By omitting the positional information of individual symbols and only considering the audible information of music symbols as output, such systems restrict themselves to replayability, as reprinting of music scores requires precise positional information [17]. While in theory semantic segmentation of the scores would go one step further and extract considerable more information - basically a classification of each pixel - two things should be noted: classifying pixels assumes that the class of each pixel is unique and mutually exclusive [18] - an assumption that might not hold for overlapping symbols but can probably be ignored for practical applications; and most traditional systems that attempt to perform semantic reconstruction operate on detected objects, not on individual pixels, thus requiring a clustering step after the semantic segmentation. Therefore we argue that detecting bounding boxes of musical objects is sufficient for performing OMR.

III. METHODOLOGY

For building a robust and extensible music object detector, we propose a machine-learning approach by using deep convolutional neural networks that operate directly on the input image. This simplifies the OMR process to the following steps: preprocessing, music object detection, and semantic reconstruction. Steps such as removing the staff lines and segmenting symbols do not need to be addressed explicitly. Existing state-of-the-art object detectors such as Faster R-CNN or R-FCN were designed to detect objects in natural scenes and often fine-tuned to work well on publicly available datasets such as COCO [20] or ImageNet [21]. Applying them out-of-the-box on a different dataset with many densely packed objects could lead to sub-optimal performance. Therefore we suggest applying a certain amount of preprocessing to the data and tailor these detectors to perform well on the task at hand.
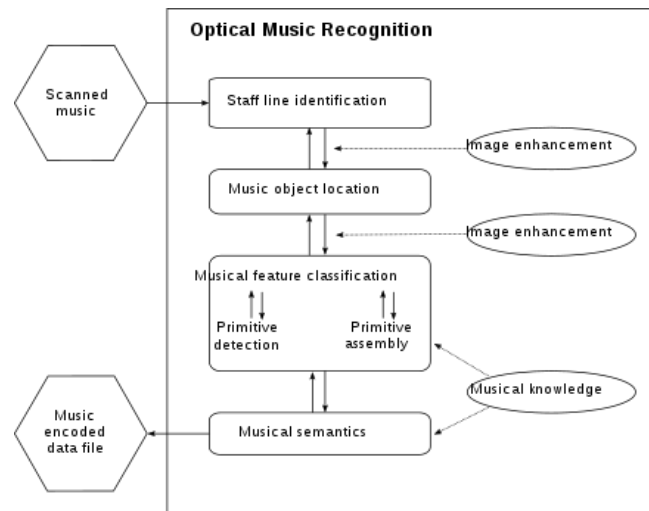
Fig3.1. System Flowchart

IV. PROPOSED WORK

1. Converts the character of ABC notation into vector, according to the idea of natural language processing (NLP) in word vector (Word Representation, Word embedding).

2. Train module: accepts vector input, output and basic parameters LSTM needed. Adopting "Back" principle, given a preceding character vector of the sequence, using the next one to present output of current input, the whole process is shown in block diagram.

3. Test module: receives vector of the test data input module and input to the LSTM module trained by step 3. Then output vector set of prediction results.

4. Converts output vector of the LSTM network test module into the character of ABC notation. It's the reverse process of step 2.

5. Converts ABC notation data into MIDI file, the output file can be played directly by music player

V. CONCLUSION

The process of recognizing music scores is typically broken down into smaller steps that are handled with specialized pattern recognition algorithms.

Many competing approaches have been proposed with most of them sharing a pipeline architecture, where each step in this pipeline performs a certain operation, such as detecting and removing staff lines before moving on to the next stage. A common problem with that approach is that errors and artifacts that were made in one stage are propagated through the system and can heavily affect the performance. For example, if the staff line detection stage fails to correctly identify the existence of the music staffs, subsequent steps will probably ignore that region of the image, leading to missing information in the output.

REFERENCES

[1] A.-J. Gallego and J. Calvo-Zaragoza, "Staff-line removal with selectional auto-encoders," Expert Systems with Applications, vol. 89, pp. 138 – 148, 2017. [2] A. Pacha and H. Eidenberger, "Towards a universal music symbol classifier," in Proceedings of the 12th IAPR International Workshop on Graphics Recognition, IAPR TC10 (Technical Committee on Graphics Recognition). New York, USA: IEEE Computer Society, 2017, pp. 35–36. [3] A. Rebelo, I. Fujinaga, F. Paszkiewicz, A. R. Marcal, C. Guedes, and J. S. Cardoso, "Optical music recognition: state-of-the-art and open issues," International Journal of Multimedia Information Retrieval, vol. 1, no. 3, pp. 173– 190, 2012. [4] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in Advances in neural information processing systems, 2015, pp. 91–99. [5] Y. Li, K. He, J. Sun et al., "R-FCN: Object detection via region-based fully convolutional networks," in Advances in Neural Information Processing Systems, 2016, pp. 379–387

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

📱 9940 572 462 🟢 6381 907 438 ✉ ijircce@gmail.com

Scan to save the contact details