INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 5, Issue 3)
Available online at: www.ijariit.com

Music staffline removal using Convolutional Neural Network

Pallavi C. L.
pallavild18@gmail.com
R V College of Enigneering, Bangalore, Karnataka

Kavitha S. N.
kavithasn@rvce.edu.in
R V College of Enigneering, Bangalore, Karnataka

ABSTRACT

For decade interest in the analysis of handwritten music, scores have been growing rapidly. It captured the focus in two type's recognition of handwritten music scores and writer identification. Many types of research have proposed different algorithms to improve recognition. Staff line removal cannot be considered as a solved problem that too when dealing with ancient music scripts. However, this work proposes to model the problem as a supervised learning classification task. This work proposes a staff line removal method using CNN to evaluate window size and mask size retaining the symbol information with improved accuracy.

Keywords— *Ancient music scripts, CNN, Supervised learning, Staffline removal, Writer identification*

1. INTRODUCTION

Image analysis and image understanding include basic image processing tasks such as segmentation and object detection. Recently many advances have been achieved in object detection using supervised learning techniques. Even though higher-level processing is more popular than lower level processing but lower level processing is an important task in many image analysis problems. These lower level tasks are used as building blocks for higher level tasks. Lower level tasks require label at pixel and super pixel level by providing by means of input and output images. In this work, supervision is provided at the pixel level and the task is modeled as image transformation.

Music is an important vehicle for cultural transmission, which is a key element as regards understanding the social, cultural and artistic trends of each period of history. A large number of music document are stored carefully for centuries in cathedrals libraries or historical archives.

In Document Image Analysis and Recognition (DIAR) analysis of music, score images are a classical area of interest. Optical Music Recognition (OMR) is used to retain the inherent music characteristics. One of the fundamental steps in OMR is staff line detection and removal. Staff lines are parallel lines that cover a large portion of music sheet and also, they are interconnected with musical symbols. Some notations are superimposed which makes recognition of music symbols more complicated. Without adequate knowledge of the location of stafflines, it is difficult to recognize, process and classify the music symbols correctly.

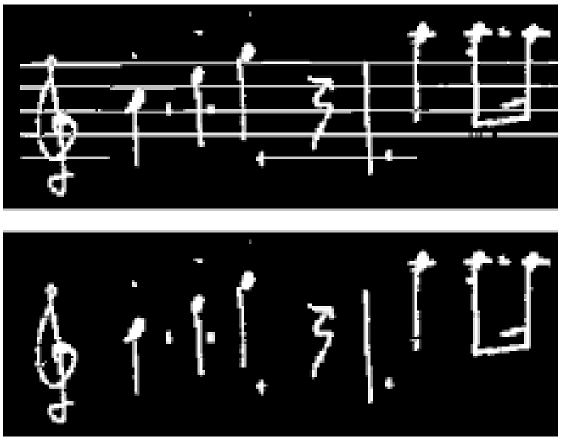

Fig. 1: Example of a binary music score image

Staff lines vary from one music document to another. Some of the difficulties found in staff removal are 1) Lines may not be parallel due to shaky handwriting 2) Lines may be curved instead of historical 3) Distortion of ink may lead to uneven thickness of lines 4) color of the staffline may have a closer resemblance to the background with symbols. There are different methods to tackle staff line removal problems. This paper is on reducing the noise as well as retaining the symbol information using CNN as a classifier model to learn the local function because they are known by their ability to learn relevant features from data. CNN's manage regularization issue by handling large input images. This paper uses knowledge of image operators for learning local functions in a small amount. Rest of the is organized as follows, in section 1.1 Related works in the field of staffline removal, in section 1.2 objectives are defined, in section 2 proposed approach of this paper, in section 3 experimentation, 4 Results and Discussions, 5 Conclusion

1.1 Related works

This paper [10] proposes a novel methodology for staff line removal, in view of Generative Adversarial Networks. Which convert staff line images into patches and feed them into a U-Net, used as Generator. The Generator means to produce staff-less images at the output.

This paper [5] proposes the use of auto-encoders, which select the fitting features of input include a set (Selection Auto-Encoders). Inside the setting of the problem at hand, the model is prepared to choose those pixels of a given picture that have a place with a musical symbol, accordingly expelling the lines of the fights.

This paper [4] proposes the utilization of Convolutional Neural Networks (CNNs) to conquer the issue on practicable window sizes. The issue of removing staff-lines in music score images is picked to assess the impacts of the window and convolutional mask sizes on the learned image operator performance

1.2 Objective

The main objective of this work is to detect/remove stafflines from the different types of stafflines obtained from handwritten music scores and to retain the symbol information by evaluating the effects of windows and mask size.

2. PROPOSED APPROACH

Dataset consists of both input and desired output images. At the time of data extraction, input images are separated from desired output images, later data is extracted from input data to form a training set. Labels are given for both input images(x) and desired output images(y).

2.1 Parameter evaluation

In object recognition, Convolutional neural networks have been used widely from the past several years. CNNs are the flexible model they are applied to various image processing tasks as they reduce the need for task-specific features. They teach filters hierarchy which extracts a high level of information in deep.

Before training the features, CNN parameters must be tuned/adjusted. 3 layers of convolutional layers, kernel size is fixed to 5 with variable window sizes with mask size of 5x5, stride 2, zero paddings. Cost entropy cost functions are used to calculate learning and dropout rates and stochastic gradient descendent algorithm is used with the epoch of 5.

2.1 Model selection

Features are extracted from each input image to form a training set. This training set is trained with CNN by extracting only white pixels from each image patch. Selected image patches are calculated for validation error using threshold frequency. If the value of error is less than the upper limit of the threshold, then those image patches are selected otherwise it is discarded.

The later output is compared with the desired output images (i. e $x_i=x_j$ but $y_i=y_j$). If the labels of the image patches match mean those patches are retained and others are discarded. Sometimes symbol information will be mistaken for staffline (i. e $x_i=x_j$ but $Y_i \neq y_j$).

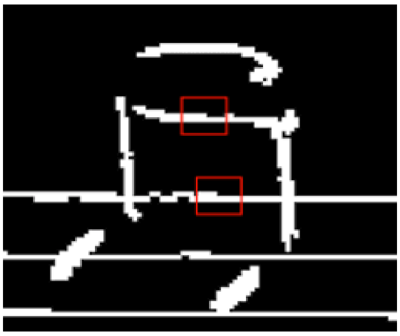

Fig. 2: Pattern of staffline and symbol look alike, there is a possibility staffline is mistaken as a symbol

3. EXPERIMENTATION

Dataset used in this experiment is ICDAR 2013 which consists of images with different types of stafflines like ideal, curvature, rotated, Kanungo, staffline thickness variation v1, v2, etc. From each type of image around 10-15 images are selected randomly as a training and test set. After selecting the best model results are stored in a separate folder. This experiment measures accuracy, specificity and recall for different sizes of windows, kernel and mask size.

4. RESULTS AND DISCUSSIONS

Window size is fixed to 256 throughout the experiment as other sizes accuracy was less. Best results are obtained from window size 256, mask size 5x5 with a threshold is fixed to 0.3. This is the best accuracy obtained when compared to other combinations of parameters.

Results from staffline type ideal, Kanungo, staffline-thickness v1 are the best one because as the thickness of the staffline increases accuracy of the results decreased and also for non-straight stafflines.

5. CONCLUSIONS

Recognition of music score images has been active research for decades where many algorithms have been proposed in order to improve recognition of music scores images. This work has shown staffline removal method using CNN which has removed staffline with improved accuracy and retaining the symbol information. It also evaluated the effectiveness of window sizes and mask sizes. Furthermore, this work can be studied to improve the results of non-straight stafflines and also for different thickness of stafflines. The study can be continued for writer identification, Converting scripts into music and also applied for different image domains.

6. REFERENCES

[1] Ana Rebelo, Ichiro Fujinaga, “Optical music recognition: state-of-the-art and open issues”, International Journal of Multimedia Information Retrieval, 2012, Volume 1, Number 3, Page 173

[2] Jorge Calvo-Zaragoza,” Music staff removal with supervised pixel classification”, International Journal on Document Analysis and Recognition archive Volume 19 Issue 3, September 2016 Pages 211-219

[3] O. Pinheiro, Pedro & Collobert, Ronan. (2015),” From image-level to pixel-level labeling with Convolutional Networks”, 1713-1721. 10.1109/CVPR.2015.7298780.

[4] S. Montagner, N. S. T. Hirata and R. Hirata, "Image Operator Learning and Applications," 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T), Sao Paulo, 2016, pp. 38-50.

[5] Antonio-Javier Gallego and Jorge Calvo-Zaragoza,” Staff-line removal with selectional auto-encoders”, journal Expert Syst. Appl., the year 2017, volume 89, pages 138-148

[6] Alicia Fornés, Anjan Dutta, “CVC-MUSCIMA: a ground truth of handwritten music score images for writer identification and staff removal”, International Journal on Document Analysis and Recognition (IJDAR), 2012, Volume 15, Number 3, Page 243

[7] T. Géraud, "A morphological method for music score staff removal," 2014 IEEE International Conference on Image Processing (ICIP), Paris, 2014, pp. 2599-2603.

[8] M. Visaniy, V. C. Kieu, A. Fornés and N. Journet, "ICDAR 2013 Music Scores Competition: Staff Removal," 2013 12th International Conference on Document Analysis and Recognition, Washington, DC, 2013, pp. 1407-1411.

[9] Calvo-Zaragoza, Jorge & J. Castellanos, Francisco & Vigliensoni, Gabriel & Fujinaga, Ichiro. (2018)." Deep Neural Networks for Document Processing of Music Score Images", Applied Sciences, 8. 654. 10.3390/app8050654.

[10] Konwer et al., "Staff line Removal using Generative Adversarial Networks," 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, 2018, pp. 1103-1108.

[11] Li Deng, "Deep Learning: Methods and Applications", Foundations and Trends in Signal Processing archive, Volume 7 Issue 3–4, June 2014, Pages 197-387

[12] Maragos, Petros. (2013)," Representations for Morphological Image Operators and Analogies with Linear Operators", Advances in Imaging and Electron Physics. 177. 45-187. 10.1016/B978-0-12-407702-7.00002-4.

[13] Anuar, Khairul & Jambek, Asral & Sulaiman, Nasri. (2016)," A study of image processing using morphological opening and closing processes", International Journal of Control Theory and Applications. 9. 15-21.

[14] S. Montagner, Igor & Hirata, Nina & Hirata Jr, Roberto. (2016)," Staff removal using image operator learning", Pattern Recognition. 63. 10.1016/j.patcog.2016.10.002.

[15] Wen, Cuihong & Rebelo, Ana & Zhang, Jing & Cardoso, Jaime. (2015)." A new Optical Music Recognition system based on Combined Neural Network", Pattern Recognition Letters. 58. 10.1016/j.patrec.2015.02.002.

[16] Ana Rebelo, Ichiro Fujinaga, "Optical music recognition: state-of-the-art and open issues", International Journal of Multimedia Information Retrieval, 2012, Volume 1, Number 3, Page 173

[17] D. Julca-Aguilar, Frank & Hirata, Nina. (2017)," Image operator learning coupled with CNN classification and its application to staff line removal", 53-58. 10.1109/ICDAR.2017.

