



HANDWRITTEN MATHEMATICAL EXPRESSION RECOGNITION AND GRADING SYSTEM

Mr. Pratik U. Patil
Department of
computer
engineering
ATS SBGI, Miraj,
Maharashtra, India

Mr. Zaid Z.
Subhedar
Department of
computer
engineering
ATS SBGI, Miraj,
Maharashtra, India

Ms. Kajol N. Khot
Department of
computer
engineering
ATS SBGI, Miraj,
Maharashtra, India

Mr. Sikandar N.
Mujawar
Department of
computer
engineering
ATS SBGI, Miraj,
Maharashtra, India

Mrs. Seema G.
Bavachkar
Department of
computer
engineering
ATS SBGI, Miraj,
Maharashtra, India

Abstract— Handwritten Mathematical Expression recognition and grading system is a challenging task in the field of pattern recognition. A lot of researchers have already worked on Handwritten Mathematical Expression recognition and have used various classifiers. In past, Convolutional Neural Network, also called CNN, has been highly used for recognizing patterns. In this paper, We propose an idea to recognize HME and evaluate offline using CNN for classification. The steps included are, first the worksheet is scanned and is sent to the work-spaces detection module where it will return all the rectangular work-spaces from the given worksheet, then the work-spaces are sent to the line extraction module to extract all the lines. The extracted lines are then passed to the character segmentation module where it will segment the character and then characters will be classified using deep learning model DCCNN. Finally, the evaluation module will assess the line and draw a green/red bounding box depending on whether the line is correct or not.

Keywords— Handwritten Mathematical Expression Recognition and Grading System, Convolutional Neural Network, Handwritten Mathematical Symbols, Handwritten Mathematical Equation Recognition

I. INTRODUCTION

Mathematics is one of the oldest studied fields and it is widely used in engineering, science, and many other fields. Converting handwritten text into machine-readable text has grown in demand. There are many researchers working on it already. Handwritten Mathematical Expression Recognition is recognizing the Handwritten Mathematical Expression(HME) that are handwritten on paper, paper is scanned, image is passed to the recognition system.

Handwritten Mathematical Expression (HME) recognition is of two types offline and online. Offline is when an image of handwritten data is taken and passed to the recognition system. Online is when handwritten data is written on any devices eg: Mobile or Tablet and this is recorded as strokes that are then passed to the recognition system[1]. In past years, computer vision has grown and improved a lot and there are many fields using it. The performance achieved by Convolutional Neural Network (CNN) is very high in terms of Handwritten Mathematical Symbol (HMS) recognition with 87.72% [2]. Even though such high performance is achieved yet there are some difficulties faced like recognizing characters such as (1 ! |) can be recognized as anyone and not the correct one. This is due to variations in the writers' style of writing.

II. RELATED STUDIES

Many research scholars have worked on recognizing offline mathematical expressions using various classifications, feature extraction, segmentation, and preprocessing methods.

Lyzandra D'souza and Maruska Mascarenhas recognized offline HME using CNN for classification. The preprocessing methods like Noise reduction, binarization, and thinning are done on the HMS dataset. The segmentation of the HME into individual math symbols is done using: recursive projection profiling algorithm. This algorithm segments the HME into individual symbols. Then merging together the small segments as one symbol like (=, i, !) is done[1]. They used a classifier called SpNet-CNN which is CNN with 83 different classes[1].

Manisha Bharambe recognized logical expressions using the Support Vector Machine(SVM) classifier with 93.8% accuracy. The preprocessing method called median filter is used to remove noise from image. In segmentation method called bounding box is used to segment each character from



expression. Feature extraction techniques such as Normalized Chain Code and Moment Invariant Features are used[7].

III. APPROACH

The process is divided into two steps: Workspace detection module and the Analysis Module.

A. Workspace Detection –

The workspace Detection module detects multiple workspaces in a given sheet of paper. It assumes that there are valid rectangular boxes in the given scanned worksheet. OpenCV is used for workspace detection. We will initially track down the rectangular boxes, then, at that point sort them depending on their positions in the worksheet. As there are many rectangles in the worksheet we only select the valid rectangles as workspace. There are many ways to find valid rectangles in a worksheet but here we choose the top 3 rectangles with maximum area.

B. Analysis Module –

The analysis module first detects and localizes characters in lines in any given single workspace, and mathematically analyzes them, and then draws red, green boxes based on whether they are correct or not. The analysis Module is divided into 3 parts Line detection, character segmentation, and evaluation.

1. Line Detection –

Our line detection module considers that there is enough gap between lines and there is some intersection between these lines and exponential characters. “First, the detected work-spaces are converted to binary images then compressed in a single array to take the forward derivative. Any place there is a line there will be a change in the derivative”[12].

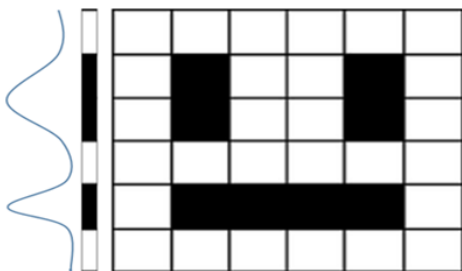


Fig. 1. Change in derivatives of binary image[12]

2. Character Segmentation and Exponential detection –

After detecting all the lines we send the extracted line images to the text_segment function where characters are segmented and sorted using OpenCV’s find contours and function sort_contours respectively. Assuming that the exponent is in the top half of the equation. We draw a horizontal baseline through the middle of the image. Any character above baseline is considered as an exponent.

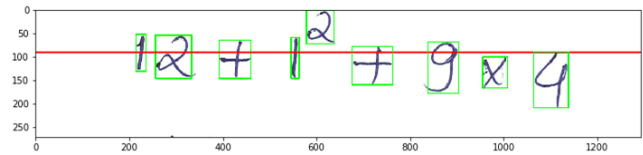


Fig. 6. Exponent Detection

Optical Character Recognition –

We use MNIST (Modified National Institute of Standards and Technology database) dataset for digits (28*28 pixels) and Kaggle’s Handwritten Mathematical symbols dataset for symbols (45*45 pixels) to train the model[12]. For classification, we use Deep Columnar Convolutional Neural Network also called DCCNN. This model achieved at most 96 % accuracy.

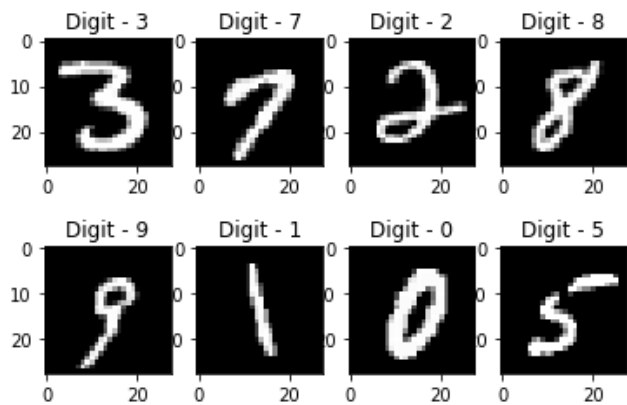


Fig. 2. MNIST Images [12]

3. Evaluation and Drawing Boxes –

In Evaluation first, we solve the given math question and save the answer. To solve the equation we use the python eval method. Then solve each handwritten line and compare its derived value with the answer stored and lastly draw a green box around the equation if the line is correct and red if the line is wrong.

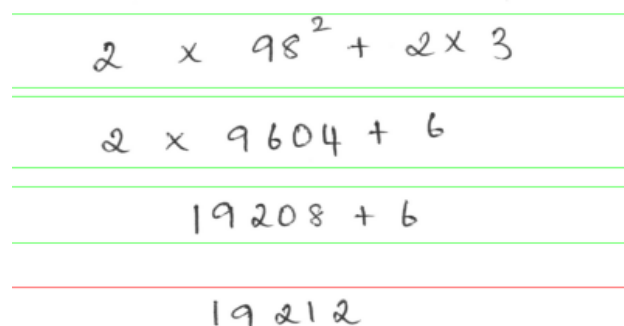


Fig.3. Sample Workspace Output



IV. CONCLUSION

We have proposed a system that will recognize offline HME and evaluate them. A handwritten mathematical expression recognition and grading system not only helps the teachers but also creates a comfortable learning experience for the students. This solution can be made even better by recognizing more complex mathematical equations like differential integral equations and trigonometry, recognizing chemical equations and recognition of cursive handwriting where the character is not separated.

V. REFERENCE

- [1] Lyzandra D'souza and Maruska Mascarenhas "Offline Handwritten Mathematical Expression Recognition using Convolutional Neural Network" 2018 International Conference on Information, Communication, Engineering and Technology (ICICET).
- [2] Irwansyah Ramadhan, Bedy Purnama, Said Al Faraby, "Convolutional Neural Networks Applied to Handwritten Mathematical Symbols Classification" 4th International Conference on Information and Communication Technologies (ICoICT), 2015 IEEE
- [3] Sagar Shinde, Rajendra Waghulade, "Handwritten Mathematical Expression Recognition using Back Propagation Artificial Neural Network" Communication on Applied Electronics (CAE), Foundation of Computer Science FCS, New York, USA Vol.4, No. 7, March 2016, pp. 1-6.
- [4] Ahmad-Montaser Awal, Harold Mouchère and Christian ViardGaudin, "The problem of Handwritten Mathematical Expression Recognition" 2010 IEEE 12th International Conference on Frontiers in Handwriting Recognition, pp. 646-651
- [5] R.Padmapriya, S. Karpagavalli, "Offline Handwritten Mathematical Expression Recognition" International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 1, January 2016, pp. 52-59.
- [6] Yassine Chajri, Abdelkrim Maarir, Belaid Bouikhalene, "A comparative study of Handwritten Mathematical Symbols recognition" IEEE 13th International Conference Computer Graphics, Imaging and Visualization, 2016, pp. 448-451.
- [7] Manisha Bharambe "Recognition of Offline Handwritten Mathematical Expressions" International Journal of Computer Applications, National conference on Digital Image and Signal Processing, DISP 2015, pp. 35-39
- [8] Abdalla Mohamed Hambal, Dr. Zhijun Pei, Faustini Libent Ishabailu, "Image Noise Reduction and Filtering Techniques" International Journal of Science and Research (IJSR) Volume 6 Issue 3, March 2017 pp. 2033-2038.
- [9] Jyotsna, Shivani Chauhan, Ekta Sharma, Amit Doegar, "Binarization Techniques for Degraded Document Images - a review" 2016 5th International Conference on

Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions), 7-9 Sep. 2016, pp163-166.

- [10] Syrine Ben Driss, M Soua, Rostom Kachouri, Mohamed Akil, "A comparison study between MLP and Convolutional Neural Network models for character recognition" SPIE Conference on Real-Time Image and Video Processing, Apr 2017, Anaheim, CA, United States.
- [11] Xue-Dong Tian, Hai-Yan Li, Xin-Fu Li and Li-Ping Zhang, "Research on Symbol Recognition for Mathematical Expressions" 1st International Conference on Innovative Computing, Information and Control (ICICIC'06), 2006 IEEE, Sept. 2006
- [12] Divyaprabha M "Computer Vision — Auto grading Handwritten Mathematical Answer Sheets" Towards data science, December 2019.