Automated Grading for Handwritten Answer Sheets using Convolutional Neural Networks

Conference Paper · October 2019

DOI: 10.1109/ICTCS.2019.8923092

Eman Shaikh
American University of Sharjah

Iman Ahmed Mohiuddin
Prince Mohammad bin Fahd University

Ayisha Manzoor
United Arab Emirates University

Ghazanfar Latif
Thompson Rivers University

Automated Grading for Handwritten Answer Sheets using Convolutional Neural Networks

Eman Shaikh, Iman Mohiuddin, Ayisha Manzoor
Department of Computer Engineering,
Prince Mohammad bin Fahd University,
Al Khobar, Saudi Arabia.

Ghazanfar Latif *, Nazeeruddin Mohammad
Department of Computer Science,
Prince Mohammad bin Fahd University,
Al Khobar, Saudi Arabia.
*Email: glatif@pmu.edu.sa

***Abstract*—Optical Character Recognition (OCR) is an extensive research field in image processing and pattern recognition. Traditional character recognition methods cannot distinguish a character or a word from a scanned image. This paper proposes a system, which is to develop a method that uses a personal computer, a portable scanner and an application program that would automatically correct the handwritten answer sheets. For handwritten character recognition, the scanned images are fed through a machine learning classifier known as the Convolutional Neural Network (CNN). Two CNN models were proposed and trained on 250 images that were collected from students at Prince Mohammad Bin Fahd University. The proposed system will finally output the final score of the student by comparing each classified answer with the correct answer. The experimental results exhibited that the proposed system performed a high testing accuracy of 92.86%. The system can be used by the instructors in several educational institutions to automatically grade the handwritten answer sheets of students effectively.**

Keywords—Handwritten Numerals Recognition, Convolutional Neural Network, Handwritten Character Recognition, Scanned document Segmentation

I. Introduction

In recent years, handwritten recognition is considered to be the uttermost engrossing and demanding analysis range in the sphere of image processing and pattern recognition. Handwritten recognition systems remarkably administer to the development of automated procedures and enhance the alliance among human and computerized systems in several operations. Nowadays, there are various technological approaches in organizations and institutions that help to reduce the time consumed for grading answer sheets manually. This is achieved by raising the accuracy and avoiding the inaccuracies caused by humans. Hence, the comparison of answer sheets with their answer keys and grading the student answers monotonously is a trivial and arduous task that must be automated.

For this purpose, Optical Character Recognition (OCR) is implemented to transform handwritten or typed text images that are captured with the help of a scanner into an electronic or machine-based text image. Predominantly, handwritten recognition systems are characterized into two categories namely, offline and online recognition. In the offline handwritten recognition systems, the handwriting written on the paper is normally apprehended by a scanner which recognizes the characters and then the completed handwritten text is obtainable as an image. Whereas, in the online recognition systems, the characters are typed from some input devices. The offline character recognition is more complicated than the online character recognition systems as writing styles may differ from one user to another and an enormous noise occurs in the offline characters during the writing of the text and scanning of the document [4, 16]. Hence, the offline handwritten recognition mechanism extends to be an effective field for research towards exploring the innovative procedures that would enhance the accuracy of handwritten recognition systems.

This paper proposes an automated system for grading handwritten answer sheets with the help of Convolutional Neural Networks (CNN). All the answer sheets were scanned separately through a portable scanner, and the scanned images were stored as black and white images. After scanning each answer sheet, the scanned images were given as an input to the segmentation algorithm that performed segmentation. This is done to separate the questions from the answers written in each box. The segmentation procedure divided the images into more comprehensive divisions and procured more relevant data. Each segmented character and digit answers were extracted to generate parameters for testing and training. The data obtained was a handwritten data set consisting of few English alphabets and numerals. The dataset was then used to score the student's answer sheet. The recognition of the student's answer was done using two CNN proposed architecture.

The remaining paper is organized as follows: Section II introduces the literature review, Section III describes the proposed framework, Section IV demonstrates the experimental results, and Section V discusses the conclusion.

II. Literature Review

The concept of handwritten recognition is a confined sphere of research in the discipline of pattern recognition and image processing over the past years and indeed there is a broad insistence for optical character recognition on handwritten scripts. In this section, an extensive analysis of extant works in handwritten recognition systems that depend on various machine learning techniques are proposed. Although the printed text recognition is considered as a clarified issue these days, handwritten text recognition remains as a demanding task, mainly due to the huge variation in handwriting among certain people including the size, orientation, thickness, format, and dimension of each written letter or digit. Various machine learning methods have been suggested for handwritten text recognition. This section describes different handwritten recognition approaches using machine learning classifiers such as the automated grading of handwritten answers and the recognition of handwritten alphabets and digits in various languages.

Brown [1] proposed an automated system using MNIST handwritten digit dataset to grade handwritten numerical answers of scanned student answer sheets using CNN. CNN was used to estimate the student answers and produced an accuracy of 95.6%. In [2], the authors implemented linear regression classifier to grade SAT standardized test essays instinctively by merging the character and word length features of essays. Kaggle dataset was used to automate the grading and resulted in an accuracy of 87.65%. In [3], an automated scoring system for multiple choice answers was implemented that allowed the users to print and scan all the answer sheets. The training time taken for each answer sheet was 35 seconds or 0.4 seconds. Feedback propagation neural network classifier was used for the implementation of the system and obtained an accuracy of 90%. Supic *et al.* [4] proposed an automated system to recognize handwritten alphabetical answers from an answer sheet containing multiple choice-based questions. Random forest classifier was implemented to automate the reading and was tested on a dataset sample that contained 3960 scanned answer sheets with an accuracy of 89.88%. Srihari et al. [5] described computational scoring methods for handwritten essays in reading comprehension tests. The handwritten response dataset consisted of 300 essays out of which 150 essays were used as training sets and 150 essays as testing sets. ANN classifier was used for the scoring methods and obtained an accuracy of 87.63%. Mahana *et al*. [6] designed an automated system for essay grading using Kaggle dataset that consisted of 13000 essays. Various essay features were extracted from the training set with the help of a linear regression model and obtained an accuracy of 91.85%.

Saengtongsrikamon *et al.* [7] developed an Optical Mark Recognition (OMR) software that was used as an OMR machine with neural networks and was then implemented in a scanner. This software captured and scored the answers of multiple choices questions with an accuracy of 95.24%. The OMR machine scanned 1,000 answer sheers using multiple scanners with varying resolutions. In [8], the authors proposed a novel local feature extraction method that was used to design a multi-language handwritten numeral recognition system. The databases used were MADBase (Arabic), MNIST (English), HODA (Persian), PMU-UD (Urdu, a database was created for this language), ICDAR (Bengali) and DHCD (Devanagri). Moreover, the same authors in [9] enhanced the feature set which was tested with the help of different classifier methods and it was found that the Random Forrest classifier achieved the best results with an average recognition rate of 96.73%.

In [10], an image processing method was proposed to automatically grade answer sheets that contained multiple choice questions. This approach enabled every user to print whatever answer sheet they wished to print and after the printing process, they utilized a normal scanner and computer to assess the answer sheets. The proposed method consumed a training time of 1.4 seconds per sheet out of the total 1000 answer sheets and detected 100% accuracy. Muangprathub *et al.* [11] presented a method to automatically grade scanned multiple-choice answer sheets using k-nearest neighbors. 560 answer sheets were evaluated in total. The proposed system operated almost three times quicker than the manual approach and the result was an average of 100% accuracy in case of nearly complete markings whereas the accuracy for the cases of incomplete markings, such as small markings, overflow, and deleted or unclean markings was obtained to be 62.42%, 93.16%, and 99.57% respectively. Patole *et al.* In [12] proposed an innovative idea for grading multiple-choice tests using a scanner that could grade a multiple-choice exam. This project used C# language to combine the computing power of C++ to evaluate each student's academic performance as well for student's to provide feedbacks on staff members. Lastly, the system was able to provide benefits which are better scalability and suitability to asynchronous mode of evaluation as compared to traditional evaluation systems. Tayana *et al.* [13] suggested a method for correcting multiple choice-based answer sheets with the help of mathematical format and k-nearest neighbors (k-NN). The database used for the manipulation of the images contained 680 certified answer sheets and 10 basic image folders containing 26 questions with four choices for each question. An overall accuracy rate of 99.85% was obtained. In [14], the author proposed an approach to grade a specifically designed multiple choice question paper with ten questions and five choices. The system could obtain 82.44% accuracy.

Ciresan *et al.* [15] proposed a handwritten character classification using CNN. Along with the MNIST dataset, a special and more challenging dataset called as the NIST SD 19 dataset was used for this purpose. CNN's were trained for around 900 epochs. The total training time consumed was twelve hours and the accuracy resulted to be 75.66%. Latif et al. [16] proposed a deep learning architecture by using Deep Convolutional Neural Networks for the recognition of Multilanguage handwritten numerals. The databases used to test the accuracy of the proposed method were MADBase (Arabic), MNIST (English), HODA (Persian), PMUdb (Urdu, a database was created as there was no pre-existing database available for this language), and DHCD (Devanagri). And the resultant overall accuracy obtained for each language was 99.322%. Singh. N. [17] suggested an active method for handwritten Devanagari characters based on ANN classifier and resulted in an accuracy of 98.65%. The training time taken for the handwritten Devanagari character recognition of 400 samples was 2.33 seconds. Kumar *et al.* [18] suggested English handwritten character recognition with the help of Kernel-based SVM and MLP Neural Network Classifiers. An isolated handwritten character dataset written by different people were considered to be the dataset. 27 features were obtained from each character during training, and these features were consumed for training the SVM with 80.96% accuracy. Rao *et al.* [19] proposed an analysis of English handwritten character recognition algorithm based on CNN. MNIST and SVHN datasets were used for this purpose. Accuracies produced by MNIST and SVHN Dataset on handwritten character recognition were 94.65% and 95.1% respectively. Jong *et al*.

[20] proposed a research on handwritten English alphabet recognition systems based on extreme learning machine. Extreme Learning Machine (ELM) classifier was used for this purpose and OCR datasets were used to train and test data. The total training time taken for each English alphabet was 0.398 seconds and the average accuracy produced was 95.513%.

In [21] an algorithm based on local adaptive thresholding and geometric features was proposed to segment different regions from scanned Arabic documents based on the Physical Layout Analysis (PLA). This method was applied to a random dataset if images from various publishers containing the text zone, image zone, and graphic zone. This algorithm achieved an average recognition of 86.71% for Text and Image block regions.

III. Proposed System

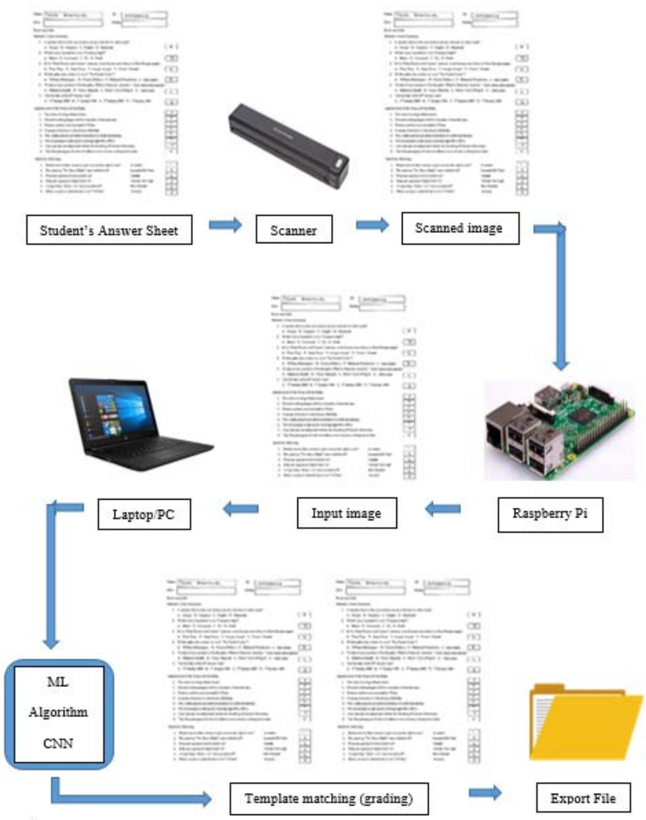


Fig. 1. Proposed System

A portable system was proposed as is illustrated in Fig. 1. It was designed to automatically recognize and grade handwritten answer sheets. For which a portable scanner 'Fujitsu Scansnap ix100' is used to scan and store the student's handwritten onto the Raspberry Pi where the scanned images are converted from pixmap to jpg format. The jpg formatted image is sent to a laptop/PC from the Raspberry Pi wirelessly which is done by using the SSH server. The laptop on which the scanned answer sheets are uploaded contains the CNN models as well as the MATLAB Segmentation code. First, the scanned images are segmented so that only the handwritten alphabets and numbers are fed to the machine learning algorithm. Thus, after segmentation, the training data is fed to the CNN models so that they are trained well to recognize the handwritten answers whereas the testing data is provided to the python scoring code which loads the trained CNN model and outputs the score for each student.

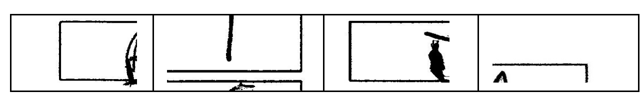
A. *Handwritten Experimental Data*

Table I depicts a comprehensive description of the segmented dataset used in the project. The dataset was collected by disturbing the template to 250 students in Prince Mohammad Bin Fahd University. As illustrated in Fig. 2, each answer sheet consisted of 20 questions, therefore a total of 5080 segmented images was obtained for 250 answer sheets. Since some of the images were not segmented properly. Only 4871 segmented images were selected (shown in Table II). Therefore, a total of 209 images was discarded as some of the scanned images captured were slightly tilted in orientation. This posed a problem because the segmentation algorithm considered each pixel value of the template. A slight difference in the orientation of the template causes different pixel values which would further yield undesirable results. Moreover, the data split chosen used the 80/20 rule, in which 80% of the dataset, i.e., 3507 images were used for training and 20% of the segmented images, i.e., 973 images were used for testing purposes.

TABLE I. COLLECTED DATASET

Answers Classes	Dataset Size
A	461
B	403
C	330
D	226
F	835
T	1090
1	271
2	251
3	251
4	249
5	252
6	252
Total	**4871**

TABLE II. INCORRECTLY SEGMENTED IMAGES



B. *Segmentation*

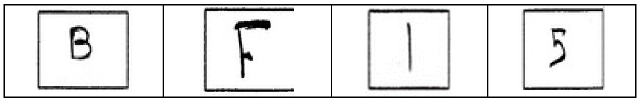
Algorithm for Segmentation of Isolated Answers from the Template is as follows:

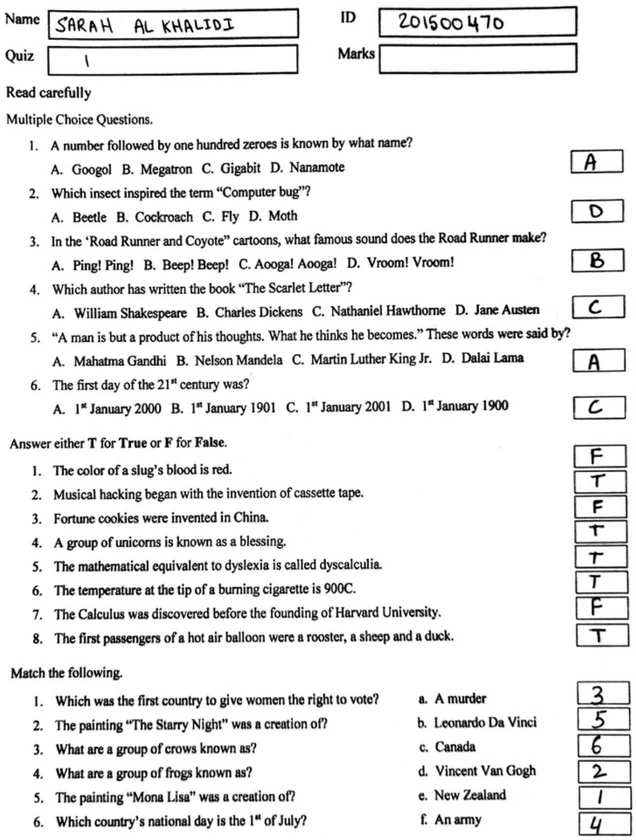
Input: *Input scanned Answer Sheet*
Output: *Isolated Answers segments with labels*
Step 1: *Start*
Step 2: *Convert input scanned document to black and white.*
Step 3: *Map the scanned document to the pre-defined template of answer sheet.*
Step 4: *Extract the questions from the input by cropping the image segments based on the template.*
Step 5: *Isolation of the twenty answers from the template is obtained by discarding the outliers.*
Step 6: *Label all the images segments based on the prior scanned template.*

Step 7: Save each segment answer in local drive.
Step 8: End

Table III depicts images of some properly segmented images from a student's answer sheet (refer Fig. 2 for the sample template).

TABLE III. SAMPLES OF SEGMENTED ANSWERS




Name SARAH AL KHALIDI
ID 201500470
Quiz 1
Marks

Read carefully

Multiple Choice Questions.

1. A number followed by one hundred zeroes is known by what name? — A
   A. Googol B. Megatron C. Gigabit D. Nanamote
2. Which insect inspired the term "Computer bug"? — D
   A. Beetle B. Cockroach C. Fly D. Moth
3. In the 'Road Runner and Coyote" cartoons, what famous sound does the Road Runner make? — B
   A. Ping! Ping! B. Beep! Beep! C. Aooga! Aooga! D. Vroom! Vroom!
4. Which author has written the book "The Scarlet Letter"? — C
   A. William Shakespeare B. Charles Dickens C. Nathaniel Hawthorne D. Jane Austen
5. "A man is but a product of his thoughts. What he thinks he becomes." These words were said by? — A
   A. Mahatma Gandhi B. Nelson Mandela C. Martin Luther King Jr. D. Dalai Lama
6. The first day of the 21st century was? — C
   A. 1st January 2000 B. 1st January 1901 C. 1st January 2001 D. 1st January 1900

Answer either **T** for **True** or **F** for **False**.

1. The color of a slug's blood is red. — F
2. Musical hacking began with the invention of cassette tape. — T
3. Fortune cookies were invented in China. — F
4. A group of unicorns is known as a blessing. — T
5. The mathematical equivalent to dyslexia is called dyscalculia. — T
6. The temperature at the tip of a burning cigarette is 900C. — T
7. The Calculus was discovered before the founding of Harvard University. — F
8. The first passengers of a hot air balloon were a rooster, a sheep and a duck. — T

Match the following.

1.	Which was the first country to give women the right to vote?	a. A murder	3
2.	The painting "The Starry Night" was a creation of?	b. Leonardo Da Vinci	5
3.	What are a group of crows known as?	c. Canada	6
4.	What are a group of frogs known as?	d. Vincent Van Gogh	2
5.	The painting "Mona Lisa" was a creation of?	e. New Zealand	1
6.	Which country's national day is the 1st of July?	f. An army	4

Fig. 2. Sample Answer Sheet

C. Convolutional Neural Network (CNN)

Parameters are an essential part of Convolutional Neural Network that helps to optimize the quality of the neural network. Their role is to avoid the overfitting and underfitting of the model for a given dataset. Changes in the parameters helps to get the desired results for a specific problem. This section talks about the different CNN parameters which were implemented to design the CNN models.

Firstly, a kernel size of more than 5 x 5 is not used since large kernel size results in a slower training time. Secondly, to minimize the error on the training data, the number of rounds of optimization that were implemented during training is increased. However, this can lead to an overfitting in the neural network which will thereby result in performance degradation during the testing phase. In order to analyze this, monitoring of error performance is done separately on the testing data as the number of epochs increases. Larger batch size requires larger memory size. Higher batch size significantly degrades the quality of the model, as measured by its ability to generalize.

The output size of an image that is produced from the hidden layer is fed into a logistic function like softmax. ReLU layer or the activation function performs an element-wise activation function max (0, x) that changes the negative values to zeros. The layer does not alter the size of the volume since there are no hyperparameters present. The softmax layer outputs a probability distribution, that is, the values of the output sum equal to 1. In addition, the softmax layer is a soft version of the max-output layer and hence it is differentiable and also resilient to outliers. Max pooling is the most used type of pooling which only takes the most important part of the input volume and the largest element from the rectified feature map. Dropout is a layer whose function is to drop out a random set of activations in the layer by setting them to zero. Moreover, it forces the network to be redundant by providing the network with the right classification or output for a specific example even if few of the activations are dropped out. It also assures that the network is not getting overfitted to the training data.

Dense is a non-linear activation function that first performs classification to the features that are extracted by the convolutional layers, then it downsamples the pooling layers. Each node present in this layer is connected to every node in the preceding layer. Adam is an optimization algorithm which is used in replacement of the classical stochastic gradient descent procedure to update the network weights iterative based on training data. It usually is a combination of RMSprop and Stochastic Gradient Descent with a momentum that uses the squared gradients to scale the learning rate like RMSprop. The cross-entropy loss function calculates the error rate between the expected value and the original value. Minimizing cross-entropy loss function approximately will help gain better performance. In the hidden/convolutional layer, all the artificial neurons are attached to the neurons of the preceding layers, in order to give out an output by picking up a set of weighted inputs. It is necessary to minimize the number of hidden layers, due to the fact that a large number of hidden layers would result in an overfit and enlarged computation.

D. Proposed CNN Models

Fig. 3 illustrates the proposed CNN architecture of Model 1. The input image for the CNN model used is of size 64x64 and then it is passed through a convolution layer of 64 filters with a kernel size of 5x5 and ReLU activation function. It is then followed by the 2 x 2 MAX Pooling layer that downsamples the image and aids in identifying the most important features. This leads to a decrease in the size of the image. Then it passed through another convolution layer of 48 filters of kernel size 3 x 3 and ReLU activation function. In order to avoid overfitting, the images then go through 20% regularization in the first dropout layer. The image further goes to more convolution, max pooling, ReLU function, and dropout layers until the sample data is ultimately converted into one-dimensional vector which happens due to the flatten layer. The final layers comprise of three dense layers that consist of 512, 256 and 12 features. The first two dense layer uses ReLU and the third dense layer uses softmax activation function which helps to convert the output into a probability distribution. The image is then recognized based on its probability distribution value.

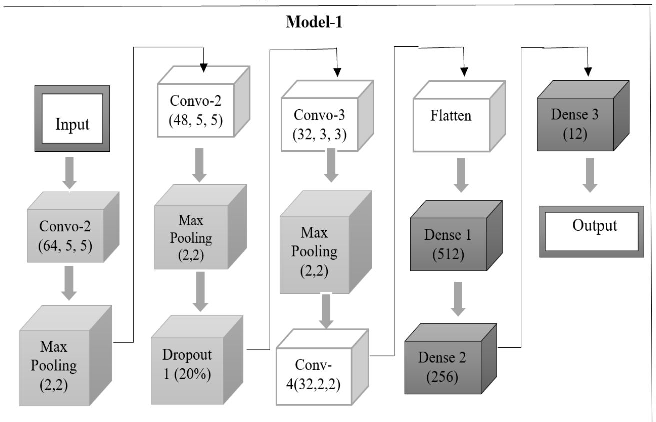


Fig. 3. Proposed CNN architecture of Model – 1

Fig. 4 illustrates the proposed Model 2 with slight changes. The first convolution layer consists of 32 filters with a kernel size of 5x5 and ReLU activation function. It is then followed by another convolutional layer of 64 filters with the same kernel size. Then the images go through 10% regularization in the first dropout layer. A 2 x 2 MAX Pooling layer is then applied to the images. Later it is passed through another convolution layer of 32 filters of kernel size 3 x 3 and ReLU activation function. In order to avoid overfitting, the images then go through 20% regularization in the first dropout layer. Further the, images undergo more convolution, max pooling, ReLU function, and dropout layers. The final layers comprise of four dense layers that consist of 512, 256, 64 and 12 features. The first three dense layer uses ReLU and the fourth dense layer uses softmax activation function.

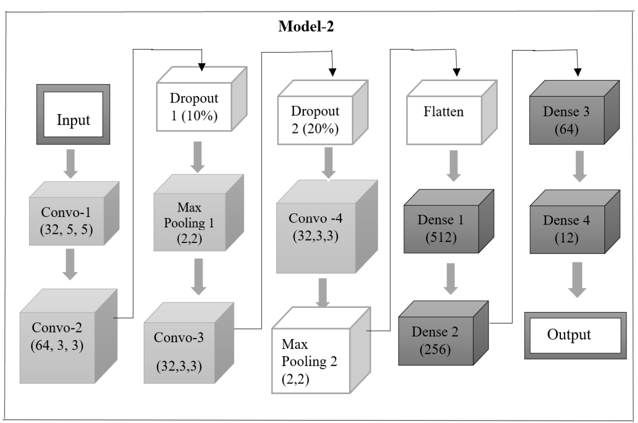


Fig. 4. Proposed CNN architecture of Model - 2

E. Scoring

Algorithm for scoring each Student's Answer Sheet is as follows:

Input*: Input Template Filled in A4 sheet*
Output*: Score of the Input Template*
Step 1: Start
Step 2: Load the CNN Model
Step 3: Obtain segmented student's answer sheets
Step 4: Read each segmented student's answer files
Step 5: Compare student's answers with True answers
Step 6: Score the answer sheets
Step 7: Display the score
Step 8: End.

Fig. 2 illustrates a student's answer sheet which provides the correct answers to all the questions. After segmentation the answers to each question will be used as refernce to score the students answersheets.

IV. Experimental Results

Table IV illustrates the experimental results accomplished by the two proposed CNN models. Epoch size of 10, 25, 50 and 100 was employed in each experiment. Furthermore, for each epoch, batch sizes of 50,100 and 250 were implemented. Several parameters were tuned and each of their influence on the level of accuracy was evaluated. A total of 12 experiments were carried out separately for model 1 and model 2. Both models were executed with the common goal of finding an optimized architecture. The experimental results demonstrate that the test accuracies generated for both the models were heavily dependent upon the number of epochs and batch sizes. An increase in the epoch size led to an increase in the test accuracy as well. Conversely, an increase in the batch size led to a decrease in the test accuracy. Optimal test accuracy was achieved when the epoch size is equal to the batch size for both model 1 (92.866 %) and model 2 (92.3274 %). Overall, the test accuracy level achieved for model 1 is better in terms of lower computation time and a minor increase in test accuracy. The result generated provided an enhanced basis for utilizing CNN architecture in the use of handwritten character recognition as a resolution to the challenges that are caused by traditional methods.

TABLE IV. EXPERIMENTAL RESULTS OF PROPOSED CNN MODEL

Model #	Batch Size	Epochs	Testing Accuracy	Computation Time (Seconds)
1	50	10	90.120 %	1772.201
1	100	10	86.451 %	1409.258
1	200	10	80.164 %	954.4838
1	50	25	91.916 %	2284.184
1	100	25	92.430 %	3919.011
1	200	25	92.353 %	2258.307
1	**50**	**50**	**92.866 %**	**4750.766**
1	100	50	92.763 %	4483.056
1	200	50	92.738 %	5266.739
1	50	100	92.840 %	10790.350
1	100	100	92.763 %	10123.380
1	200	100	92.840 %	9349.122
2	50	10	90.4799 %	1803.066
2	100	10	90.4799 %	1363.192
2	200	10	81.011 %	1885.508
2	50	25	92.2504 %	3579.950
2	100	25	92.1991 %	3295.304
2	200	25	90.8648 %	4753.194
2	50	50	92.3274 %	6125.266
2	100	50	92.3531 %	6142.417
2	200	50	92.3531 %	5819.936
2	50	100	92.4301 %	11866.290
2	100	100	92.4044 %	12197.170
2	200	100	92.1735 %	11079.200

V. Conclusion

Offline handwritten recognition systems based on machine learning algorithm has significant importance in the research field. However, it is a difficult recognition due to the presence of odd characters or similarity in shapes for multiple characters. The paper proposed a system that was implemented to recognize the handwritten characters and then display the final score of the student. The system was evaluated from a dataset that consisted of 250 answer sheets and this data was tested by using two deep convolutional neural network models. The results attained a high accuracy with 92.86% testing accuracy. The accuracy of the system was less as compared to the ones mentioned in section II as the system used its own handwritten data set. In future work, the segmentation algorithm can be improved to attain a higher percentage of accuracy for segmentation of the images. Moreover, the proposed CNN architecture can also be enhanced to achieve much higher performance and accuracy in displaying the score of the student.

References

[1] Brown, M. T. (2017). Automated Grading of Handwritten Numerical Answers. In *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)* (pp. 279-284). IEEE.

[2] Murray, K. W., & Orii, N. (2012). Automatic essay scoring. *IEICE Transactions on Information and Systems*, *102*(1), 147-155.

[3] Alomran, M., & Chia, D. (2018). Automated Scoring System for Multiple Choice Test with Quick Feedback. *International Journal of Information and Education Technology*, *8*(8).

[4] Cupic, M., Brkic, K., Hrkac, T., Mihajlovic, Z., & Kalafatic, Z. (2014, May). Automatic recognition of handwritten corrections for multiple-choice exam answer sheets. In *Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on* (pp. 1136-1141). IEEE.

[5] Srihari, S., Collins, J., Srihari, R., Srinivasan, H., Shetty, S., & Brutt-Griffler, J. (2008). Automatic scoring of short handwritten essays in reading comprehension tests. *Artificial Intelligence*, *172*(2-3), 300-324.

[6] Mahana, M., Johns, M., & Apte, A. (2012). Automated essay grading using machine learning. In *Document Analysis and Recognition, ICDAR. 10th International Conference* (pp. 1206-1210). IEEE.

[7] Saengtongsrikamon, C., Meesad, P., & Sodsee, S. (2009). Scanner-based optical mark recognition. *Information Technology Journal*, *5*(1), 69-73.

[8] Alghazo, J. M., Latif, G., Alzubaidi, L., & Elhassan, A. (2019). Multi-Language Handwritten Digits Recognition based on Novel Structural Features. *Journal of Imaging Science and Technology*, *63*(2), 20502-1.

[9] Alghazo, J. M., Latif, G., Elhassan, A., Alzubaidi, L., Al-Hmouz, A., & Al-Hmouz, R. (2017). An Online Numeral Recognition System Using Improved Structural Features–A Unified Method for Handwritten Arabic and Persian Numerals. *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)*, *9*(2-10), 33-40.

[10] Chai, D. (2016, December). Automated marking of printed multiple-choice answer sheets. In *2016 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)* (pp. 145-149). IEEE.

[11] Muangprathub, J., Shichim, O., Jaroensuk, Y., & Kajornkasirat, S. (2018). Automatic Grading of Scanned Multiple-Choice Answer Sheets.

[12] Patole, S., Pawar, A., Patel, A., Panchal, A., & Joshi, R. (2016, March). Automatic system for grading multiple choice questions and feedback analysis. *IEEE International Journal of Technical Research and Applications, 12*(39), 16-19. IEEE.

[13] Tavana, A. M., Abbasi, M., & Yousefi, A. (2016, September). Optimizing the correction of MCQ test answer sheets using digital image processing. In *2016 Eighth International Conference on Information and Knowledge Technology (IKT)*(pp. 139-143). IEEE.

[14] Abbas, A. A. (2009). An automatic system to grade multiple choice questions paper-based exams. *Journal of university of Anbar for Pure science*, *3*(1), 174-181.

[15] Ciresan, D. C., Meier, U., Gambardella, L. M., & Schmidhuber, J. (2011, September). Convolutional neural network committees for handwritten character classification. In *Document Analysis and Recognition (ICDAR), 2011 International Conference on* (pp. 1135-1139). IEEE.

[16] Latif, G., Alghazo, J., Alzubaidi, L., Naseer, M. M., & Alghazo, Y. (2018, March). Deep Convolutional Neural Network for Recognition of Unified Multi-Language Handwritten Numerals. In *2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR)* (pp. 90-95). IEEE.

[17] Singh, N. (2018, February). An Efficient Approach for Handwritten Devanagari Character Recognition based on Artificial Neural Network. In *2018 5th International Conference on Signal Processing and Integrated Networks (SPIN)* (pp. 894-897). IEEE.

[18] Kumar, P., Sharma, N., & Rana, A. (2012). Handwritten Character Recognition using Different Kernel based SVM Classifier and MLP Neural Network (A COMPARISON). *International Journal of Computer Applications, 53*(11), 413-435.

[19] Rao, Z., Zeng, C., Wu, M., Wang, Z., Zhao, N., Liu, M., & Wan, X. (2018). Research on a handwritten character recognition algorithm based on an extended nonlinear kernel residual network. *KSII Transactions on Internet & Information Systems, 12*(1), 25-31.

[20] Jeong, S. H., Nam, Y. S., & Kim, H. K. (2003, August). Non-similar candidate removal method for off-line handwritten Korean character recognition. In *Document Analysis and Recognition, 2003. Proceedings. Seventh International Conference on* (pp. 323-328). IEEE.

[21] Al-Dobais, M. A., Alrasheed, F. A. G., Latif, G., & Alzubaidi, L. (2018, March). Adoptive Thresholding and Geometric Features based Physical Layout Analysis of Scanned Arabic Books. In *2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR)* (pp. 171-176). IEEE.