Electrical-Electronics Engineering

Bartin University – Turkey March 4, 2021

Dear Editors,

We thank the editors and reviewers for their efforts, time and useful comments.

We have edited the manuscript addressing your valuable comments. We believe that your comments made the manuscript scientifically suitable to be published.

Mohammed Salemdeeb

PhD, Electronics and Communication Engineering.
Bartin University, Turkey,
Faculty of Engineering,
Electrical-Electronics Engineering department.

On behalf of all authors.

Reviewer 3 (Anonymous)

Basic reporting
A deeper error analysis is required.

More error analysis sentences are added at line 323, 324, 330 and 347.

Update at line 323,
“All of the previous datasets were divided into training and test sets by their authors where the instances in the test set were collected from a different source (different writers for CR and different photographers for fashion) from the training set’s source. The performance evaluation is done based on CNN type (stacked is simpler than spars), number of layers, number of learnable parameters and recognition error.“

Other updates are written below.

Experimental design
The authors need to discuss the evaluation procedure.

All the reported results were concentrated on the classification error, the datasets were divided into training and test sets and the evaluation is based on CNN type, classification error, number of layers and number of learning parameters.
More sentences to discuss the evaluation procedure are added at lines 324, 330 and 347.

Validity of the findings
The authors can discuss the results for other performance metrics.

We discuss the results for performance metrics classification error, number of learnable parameters, number of layers and memory usage of FDCNN. In the previous related work these are the most common used metrics and we follow them to compare our performance and introduced dataset. Some of datasets were not entered to the training process but it was used for test, Zemris dataset was not seen in the training and the model was trained on other datasets but the model was tested only on Zemris, which reflects the validity of the results. More sentences are added to discuss the validity of the results at line 324, 330 and 347.

Comments for the Author
The literature review needs to be updated with some of the recent works. Discussing the following works would make the manuscript richer for the readers:

DevNet: An Efficient CNN Architecture for Handwritten Devanagari Character Recognition. Int. J. Pattern Recognit. Artif. Intell. 34(12): 2052009:1-2052009:20 (2020)
Character recognition based on non-linear multi-projection profiles measure. Frontiers Comput. Sci. 9(5): 678-690 (2015)

Relative Positioning of Stroke-Based Clustering: a New Approach to Online Handwritten Devanagari Character Recognition. Int. J. Image Graph. 12(2) (2012)
Artistic Multi-character Script Identification Using Iterative Isotropic Dilation Algorithm. RTIP2R (3) 2018: 49-62
Character Recognition Based on DTW-Radon. ICDAR 2011: 264-268
Spatial Similarity Based Stroke Number and Order Free Clustering. ICFHR 2010: 652-657
Dtw-Radon-Based Shape Descriptor for Pattern Recognition. Int. J. Pattern Recognit. Artif. Intell. 27(3) (2013)

Thanks for these comments, we believe that:
"DevNet: An Efficient CNN Architecture for Handwritten Devanagari Character Recognition. Int. J. Pattern Recognit. Artif. Intell. 34(12): 2052009:1-2052009:20 (2020)" is recent and related to our research regarding both character recognition problem and CNN. We have updated the literature review citing recent "DevNet" reference.

Reviewer 4 (Kanika Thakur)

Basic reporting
A new approach to LP identification by stacking two CNN networks is discussed in the paper. The core of the convolution block, which is the convolution Layer is followed by batch normalization and a non-Linear activation layer.

Experimental design
There is a well-designed experimental section. The manuscript's strongest achievements are the suggested Full Depth CNN (FDCNN) model and the recommendation for a new license plate data set called LPALIC. The strength of the paper is the selection of parameters, filters, and the process of training. The manuscript also gives an overview of FDCNN with respect to the use of memory.
The drawback of the manuscript is that the results obtained during the testing process on the test dataset are missing and no validation dataset has been used, so it is unclear that how fine-tuning of the model hyperparameters has been performed and the problem of overfitting has been addressed. I think the writers should be able to update the manuscript before publishing to make all these points clear and further illustrate and solidify their already good findings

Thank you, we did a lot of experiments and the written results are the final average results. Since we have 11 tables in our paper and we see that we cannot add more tables otherwise the paper will be too long.
Most of the tests done on LPALIC and ALPR datasets were designed to overcome the overfitting problem by training the model on a dataset and testing the model on another characters dataset.

For example, to test UCSD dataset we trained the model on our Turkish and EU LP characters set. FDCNN is tested on many datasets and the average performance is very good on test sets. Fine-tuning of the model hyperparameters has been performed as described at line 284 by retraining the model again using other training algorithm and different training settings. When we did, the results get better which may be seen as a low effect of the overfitting problem.

We updated our manuscript to make all these points clear and further illustrate and solidify our findings at lines 324, 330 and 347.

Update at Line 324 “However, Khaled et al., (2010) used his dataset for both training and testing, FDCNN could classify the whole dataset (as a test set) of (Khaled et al., 2010) with error of 0.46\% whereas the training was done on characters collected and cropped manually from public KSA LP images.”

Update at line 330 “Zemris, UCSD, Snapshots and ReId datasets were not used in the training process but the proposed FDCNN was tested on each of them as a test set to ensure that the model was fitted to character features, not to a dataset itself. For UFPR dataset, FDCNN was tested two times on UFPR test set, training on only the training set of UFPR and training on both UFPR and LPALIC characters.”

Update at line 347 “In the manual split in Table 11, the country’s characters training and testing sets were used to train and test FDCNN. In trained on other countries, the FDCNN was trained on both the country’s characters training set and other countries characters but tested only on that country’s test set. In the random 80/20 split, the country’s characters were split randomly into training and testing sets, and FDCNN was trained on both the split country’s characters training set and other countries characters but tested only on that split country’s test set, a lot of random split tests were done and the average errors were reported in the table. Those different test analyses were done to validate and evaluate the results and reduce the overfitting problem.”

Validity of the findings
All outcomes are well defined and codes are given, but to ensure the validity of the data, the points listed must be answered. I assume that the relevance of the results can be more readily asserted if the issues are discussed and explained.

Comments for the Author
The paper is clear, but there are few points that require Clarification

1. Explains the methodology used in order to correctly recognize Arabic zero numbers and letters written in continuous style.

Thank you for this comment. This paper is interested in isolated character recognition. However, the Arabic words are written in a continuous style but in this study we used the isolated Arabic character datasets.

At line 66 we described the solution proposed by Abdleazeem and El-Sherif (2008), in our paper we used the same logic of size-sensitive feature by zero character half size reduction since it has a smaller size than other characters.
A new explanation sentence is added at line 303," The same logic of size-sensitive feature proposed in (Abdleazeem and El-Sherif, 2008) is used to solve the problem of Arabic zero character by half size reduction for Arabic zero character image (in MADbase dataset) since it has a smaller size than other characters."

2.Discuss the criteria based on which a character is labeled as difficult /easy as specified in line number 342.

Agreed, the criteria here is difficulty of manually labelling the character images in the dataset preparing stage (labelling by human).

A clarification parentheses is added at line 342. (difficult at manual labelling the character images in the dataset preparing stage)

3.You should clarify, why one should go for the proposed FDCNN model, when Assiri,2019 stacked method has outperformed.

At line 297, we wrote a simple clarification.
We edited the clarification paragraph at line 297 to be like this:
"However, the architecture in (Assiri, 2019) has 15 layers with 13.12M parameters while FDCNN has 12 layers with 1.69M parameters which means that FDCNN is simpler and 7 times faster (in terms of the number of parameters 13.12/1.69). The results in Assiri 2019 were obtained utilizing data augmentations (not used in FDCNN training), different training processes (FDCNN training process is simpler as described in the previous section) and Dropout layers before and after each pooling layer with different settings, but, FDCNN has no Dropout layer and showed good results on MNIST."