Offline Handwriting Recognition-Critical Analysis

Aman Bhansali
B20ME010
bhansali.1@iitj.ac.in

Indian Institute of Technology Jodhpur

This review is for the paper Offline Handwriting Recognition with Multidimensional Recurrent Neural Networks written by Alex Graves and Jurgen Schmidhuber. Alex Graves is a research scientist at Deep Mind. Jurgen Schmidhuber is a co-director of the Dalle Molle Institute for Artificial Intelligence Research. The title of the paper perfectly suits the content it contains. The authors have mentioned some efficient techniques, extended some algorithms for higher dimensions, and applied some new combinations of algorithms of machine learning so that computers recognize the handwritten characters/words/sentences more accurately.

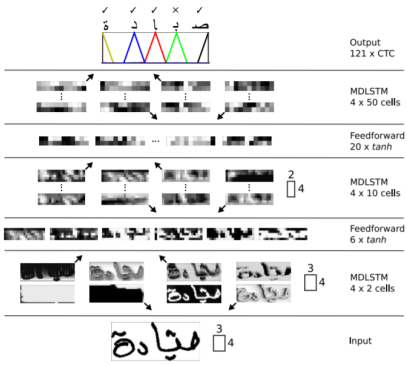
The abstract of the paper gives good information about what the paper is all about and what are the key techniques being used to solve the given problem. The resulting model that is built using the algorithms makes it possible for the recognition of both online and offline handwriting and that also it does not require any segmentation, pre-processing, or any post-processing of input or output for a particular kind of input/language. The paper has all its contents well structured which gives a proper intuition of each step on which the model would be working.

The paper discusses some issues that other algorithms face like HMM. But, HMM needs that its input features should meet some assumptions it imposes. Also, the features for this need to be redesigned for different languages which means that the same system would not be able to work/detect different languages. The methods which were imposed in this paper though perfectly solve all the above issues well.
The model uses algorithms such as multidimensional RNN which makes the model robust towards warping too. Long short-term memory (LSTM) is being used which can store much information of previous important inputs for long textual data. Thus, using this helps store all the information and can also forget the data not important after a point.

The references mentioned help understanding each of the key algorithms, the equations mentioned in the paper. Also, the paper gives justifications regarding most of the equations and techniques used ex. explains why there are L+1 units used in CTC. The references mentioned contain some papers which also contain a pseudo-code for implementing the algorithm like multi-dimensional recurrent networks and multi-directional recurrent networks. Also, the paper mentions the practical implementation of the model in a competition IFN/ ENIT database of handwritten Arabic words made which also provides the readers a self-motivation to learn from the idea and also implement the same. Also, it mentions the parameters used during the implementation and also the parameters needed to train the model. The accuracy scores gained were also greater than many other models which participated thus, outstanding the model. Some more scenarios are discussed and solutions for those are also proposed such as regarding the size of input which would not be the same always and hence block size can't be changed always.
The authors also participated in a competition ICDAR 2007 Arabic handwriting recognition competition to test this model. The goal of the competition was to identify the postcodes of Tunisian town and village names. The model outperformed and gave very good accuracy and in fact, it was greater than all the others that models participated.

	SET f			SET s		
SYSTEM	top 1	top 5	top 10	top 1	top 5	top 10
CACI-3	14.28	29.88	37.91	10.68	21.74	30.20
CACI-2	15.79	21.34	22.33	14.24	19.39	20.53
CEDAR	59.01	78.76	83.70	41.32	61.98	69.87
MITRE	61.70	81.61	85.69	49.91	70.50	76.48
UOB-ENST-1	79.10	87.69	90.21	64.97	78.39	82.20
PARIS V	80.18	91.09	92.98	64.38	78.12	82.13
ICRA	81.47	90.07	92.15	72.22	82.84	86.27
UOB-ENST-2	81.65	90.81	92.35	69.61	83.79	85.89
UOB-ENST-4	81.81	88.71	90.40	70.57	79.85	83.34
UOB-ENST-3	81.93	91.20	92.76	69.93	84.11	87.03
SIEMENS-1	82.77	92.37	93.92	68.09	81.70	85.19
MIE	83.34	91.67	93.48	68.40	80.93	83.73
SIEMENS-2	87.22	94.05	95.42	73.94	85.44	88.18
Ours	**91.43**	**96.12**	**96.75**	**78.83**	**88.00**	**91.05**

Also, there are some cons rectifying which could make our model more efficient. Each pass through the input took around 1 hour which could become even more if the model converges after many iterations. Some of the methods like the multidimensional LSTM which is of great importance for model making but are not explained completely that is the coding part like using loops, variable initialization is not discussed in the paper. The paper uses a lot of algorithms which surely is increasing the accuracy of the model but also increasing its complexity a lot. Also, the part where the model is presented working with a sample should have been explained a better way like why a particular dimensional activation layer is used which would give a better understanding of the paper also give the complete feel of the actual model's working. Equations mentioned in the



paper contains many parameters such as activation parameters and a feel of it. Also, there is no mention of a detailed reason for the time cost of the model which should be provided as it would help others to look at that particular steps and optimize them to somewhat improve the model.

The time consumption problem could be a topic for future research works too. Also, some preprocessing of input could help us get more accuracy like de-slanting could be done which would make the characters more easily recognizable. Deslanting means that if the image of the text given is tilted at any side and it just rotates the picture of the text Also, scaling of the pixel values could be done. This could be done either by using scalar or just subtracting the pixel value of the input data from 255 and dividing it by 255 or using sigmoid/tanh and then multiplying the same with the input.

References

1. https://www.cs.toronto.edu/ graves/icml$_2$006.*pdf*
2. https://arxiv.org/pdf/0705.2011.pdf