Journal of Innovative Image Processing (ISSN: 2582-4252)
www.irojournals.com/iroiip/

Deep Learning based Handwriting Recognition with Adversarial Feature Deformation and Regularization

Yasir Babiker Hamdan[1], A. Sathesh[2]

[1]International University of Africa (IUA), Khartoum, Sudan
[2]Department of Electronics and Communication Engineering, Eritrea Institute of Technology, Eritrea

E-mail: [1]yasir20ap@iua.edu.sd, [2]sathesh4you@gmail.com

Abstract

Due to the complex and irregular shapes of handwritten text, it is challenging to spot and recognize the handwritten words. In low-resource scripts, retrieval of words is a difficult and laborious task. The need for increasing the number of samples and introducing variations in the extended training datasets occur with the use of deep learning and neural network models. All possible variations and occurrences cannot be covered in an efficient manner with the use of the existing preprocessing strategies and theories. A scalable and elastic methodology for wrapping the extracted features is presented with the introduction of an adversarial feature deformation and regularization module in this paper. In the original deep learning framework, this module is introduced between the intermediate layers while training in an alternative manner. When compared to the conventional models, highly informative features are learnt in an efficient manner with the help of this setup. Extensive word datasets are used for testing the proposed model, which is built on popular frameworks available for word recognition and spotting, while enhancing them with the proposed module. While varying the training data size, the results are recorded and compared with the conventional models. Improvement in the mAP scores, word-error rate and low data regime is observed from the results of comparison.

Keywords: Handwriting recognition, deep learning, feature deformation, error estimation, neural network

1. Introduction

Every day, several innovations in the areas of artificial intelligence and deep learning, emerge with technological advancements [1]. Computer models are used for learning and execution of pattern recognition tasks directly from images, text, or sound in deep learning.

DOI: https://doi.org/10.36548/jiip.2021.4.008
Received: 04.11.2021, received in revised form: 30.11.2021, accepted: 14.12.2021, published: 24.12.2021

Deep neural networks are used in a number of active research areas, which include handwriting recognition [2]. Handwriting recognition is a simple process for humans, but it is a challenging task for computers. In the field of computer vision, images can be manipulated and processed using image processing techniques. Character recognition requires a number of steps, which include acquisition, feature extraction, classification, and recognition [3]. Handwriting recognition is a machine's capacity to receive and analyze manually written input from an external source, such as an image [4]. Handwriting Recognition can be done either in online or offline mode. In the online recognition method, the user is identified once he starts to write on electronic devices like mobiles and iPads. In the offline recognition method, the image, which consists of a user's handwriting, was given as an input in the system and the user would be identified [5, 6]. For recognition purposes in both methods, the strokes the user makes while writing the letters will be stored in the system.

Handwritten signature recognition can also be constrained with the advantages like cost-effective technology, rapid checking time, minimal intrusive threshold, and fast public acceptance when compared to other rising biometric technologies [7, 8]. In 2007 a software named CEDAR FOX was developed in the University of Buffalo to identify whether the two different documents were written by the same person or not. The Handwriting verification is done in two steps [9]. First, the test document is compared with a known sample document, to know the percent of variation done by a person while writing a document. Second, the test document is compared with a minimum of 4 documents to understand the writing habit of the user [10]. However, it was difficult to recognize the users due to the different styles. A large dataset will be needed for storing the information and making the computers automated. Thus, neural computers can be used [11]. A neural computer functions in a fundamentally different way. Neural computers are instructed to evaluate the input data and provide an appropriate output to the user. By applying these techniques, we can recognize the handwritten characters of the user.

The major contributions made in this paper are as follows:

1. Spotting and recognition of handwriting is performed through a scalable solution offered by integration of adversarial learning in low resource scripts with high dimensional convolutional feature space.

2. Even with limited amounts of data, the adversarial generator introduces several deformations that encourage learning from variations in handwriting at the task network.

3. Different baselines are used for comparison of the proposed adversarial augmentation scheme. When compared to the state of the art systems, the proposed model offers better performance in handwriting spotting and recognition.

4. Real-time handwritten data analysis is performed in an enhanced manner while improving the performance of low resource script cases also.

Domain adaptation, cross-domain image translation and several other tasks influence the successful implementation of adversarial learning [12]. Due to these factors, spatial transformations are used in a high dimensional feature space to augment word images with a paradigm based on generative adversarial learning [13]. Spotting or recognition is performed by the original network with the addition of the Adversarial Feature Deformation and Regularization Module (AFDRM). The trivial and easily learnable features can overcome overfitting with the integration of this module. Rare deformations in the real-world testing data can be generalized well with the proposed module due to its enhanced framework [14]. When difficult variations are faced, invariances are learnt by the task network through the generation of hard examples generated by the adversarial generator where the task network and adversarial generator (AFDRM) are jointly trained, and gradually improves the performance over time.

2. Literature Review

Detailed research and reviews are performed in great detail by researchers over the past regarding handwriting recognition. However, more accurate and better techniques are still developed by several researchers. Bag-of-n-grams approach and word-embeddings models are compared in [15] where the word-embeddings model offers enhanced performance. A word encoding technique based on attributes is presented in [16] using a ConvNet where existing dictionary word profiles are correlated with the words in the input image that are included in the spatial parts of n-gram for estimation of the frequency profile. Pyramidal Histogram Of Characters (PHOC) attributes are embedded with the handwritten words images whose holistic representations are predicted with the terminal fully connected layers with the help of VGG-Net [17]. Textual embedding space is embedded with features in architectures presented in [18-

20]. Searches are performed in a distributed word embedding space where regional features are encoded by an end-to-end model in the word-spotting model driven by the region proposal network presented in [21]. Several works like [22] make use of RNN training approaches based on Connectionist Temporal Classification (CTC) criterion used for sequence discriminative training are presented in [23].

Word transcriptions are computed in MDLSTM, LSTM and other recurrent networks using ConvNet by engineering a sequence of image features [24]. In order to enhance detection accuracy, before the sequence-to-sequence transcription, images are spatially reoriented with the help of an attention mechanism based on affine-transformation in [25]. For extension of the original dataset, it is essential to preprocess images in most of the techniques discussed so far. The standard datasets consisted of almost 4000 classes in character recognition techniques for Chinese handwriting [26]. Extensive and large datasets also use extended datasets with augmentation techniques. A smaller number of hard examples are considered in the datasets while enhancing the accuracy and effectiveness of the process using online hard example mining technique with different sets of approaches. Realistic synthetic data is created with the incorporation of generative modeling in several approaches in recent years with the advancements in GAN and adversarial learning for stable training of GAN based on the architectural guidelines described in [27].

3. Proposed Methodology

The Hand-Writing Spotting (HWS) and Hand-Writing Recognition (HWR) framework exhibit generic augmentation techniques which are not sufficient for identifying and generalizing real-world handwritten data. This is especially the case when low-resource scripts with small datasets are used which account for a minimal amount of observed irregularities. In this proposed work, a modular deformation network is proposed that is trained to understand the various parameters involved in deforming the features of the system, thereby enabling it to accept and adapt to various irregularities and difficult examples. Consider ‘I’ as the input image given to a task network K. Here the task network denotes word spotting or word recognition network, coupled with cross-entropy loss and CTC loss represented as task loss L_T. An Adversarial Feature Deformation Module (AFDM) is introduced in the middle of the task network in this proposed model.

The successful convolutional sections of T are denoted as T_A and T_B and dissection of T will result in three parts namely R, T_A and T_B. In the following eq (1),

$$F = T_A(I)$$

where, F denotes the output feature map of T_A and I is the image. At the time of interference, only T is used.

Figure.1. shows the architecture of the Adversarial Feature Deformation Module used with the training network. The following are the various components of the architecture:

1. A localization network A is used on θ parameters for prediction. The transformation matrix T_θ is calculated with the help of these parameters.

2. 'S' is the sampling grid generated by the Grid Generator based on the transformation on the S' grid denoting F' coordinates.

3. The original feature map and grid S are passed through Bilinear Sampler in order to get F' which is the target feature map.

4. The Thin Plate Spline (TPS) transformation is chosen to be used in AFDM due to their capacity to deform a plane elastically and its degree of flexibility.

5. The control points represent parameters which are used in Grid Generator to define transformation function representing base control points.

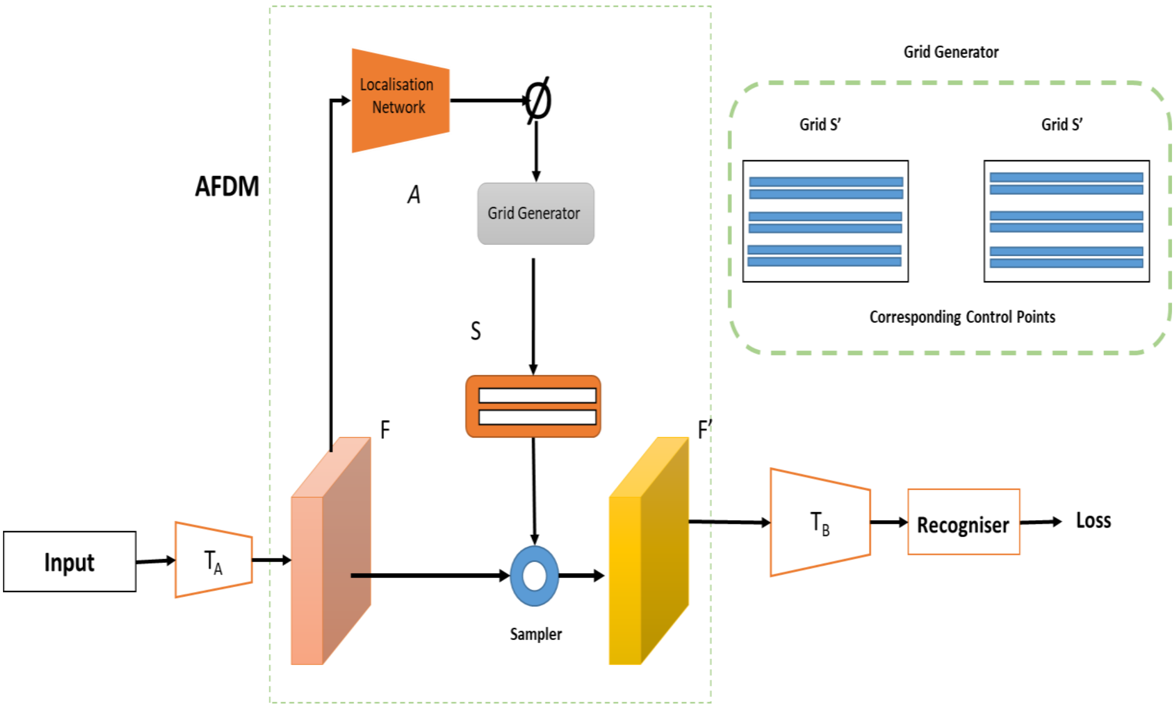

Figure 1. Architecture of the proposed work

In this proposed work, two sets of Latin Scripts are used as datasets: RIMES and IAM which are extensively used in handwriting document analysis. For HWS and HWR in Latin script IAM holds the largest datasets, paving way to projecting the effectiveness of the proposed feature wrapping methodology for various sizes of training sets as shown in Fig.2. On the other hand, for low-resource scripts, an effective demonstration of the proposed model is made with Devangari and Bangla, two Indic scripts that show the advantages of using AFDM for adversarial training. Here Bangla and Hindi are one of the popular languages that are spoken and can be used in scripts Bangla and Devangari respectively. When compared with Latin, these two scripts are said to be more complex because of the presence of complex cursive shapes and modifiers. Based on this research, only one publicly available dataset is possible that holds a total of 16,128 and 17,091 words in Devangari and Bangla. IndDEV and IndBAN are the two datasets available wherein they are used in IAM for partitioning in testing, validation and training available datasets.

4. Experiments and Results

Based on experimental analysis, it was identified that it is necessary to first pre-train the task networks for a particular set of iterations in order to study the basic model for better understanding of the various shapes of characters. On training both the networks, the AFDM was found to overpower the task network without proper understanding of the representation. Hence, the task network is initially trained for a total of 10000 iterations in the absence of AFDM. On including AFDM, the intermediate convolutional feature maps are deformed. Training takes place for 5 interactions in the network 'A'. It is further observed that because of flexibility issues, the TPS will be deformed resulting in the failure of task network.

This stability issue is addressed by a simple trick wherein only 50% of the data is deformed in a particular set of AFDM while the remaining are kept unchanged, thereby improving stability of the system. 4 convolutional layers with 3 x 3 kernel size and 2 strides are used in the localization network, predicting 20 parameter values with the help of tanh activation as shown in Fig.2. The total sub-map divisions are limited to four and the size of the batch is fixed at 32. Based on earlier initialization, both the AFDM and task network are trained using 100k iterations. AFDM and task network make use of Adam optimizer, holding the learning rate at 10-4 for network and 10-3 for localization network of AFDM. There are a total of 13 convolutional layers present along with 3 fully connected layers and SPP layers.

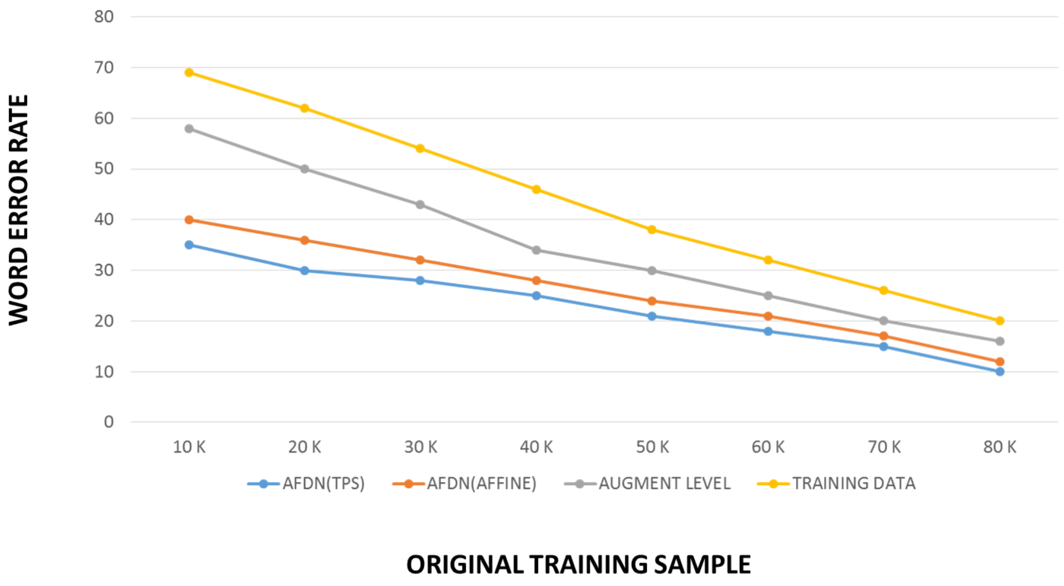

Figure 2. Word Error Rate for unconstrained HWR

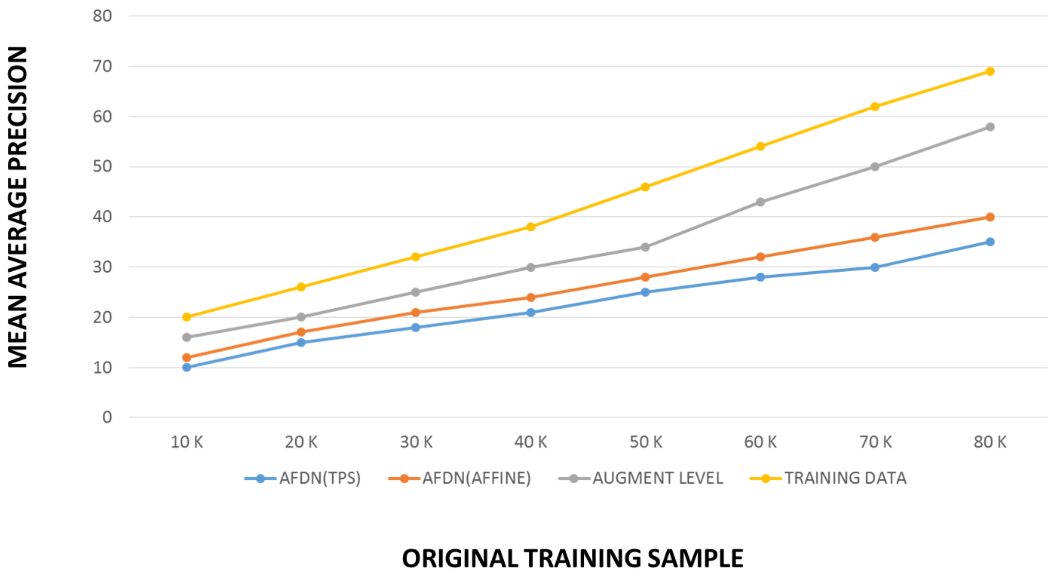

Figure 3. Mean average precision observed during training

The improvements observed are recorded for various augmentation techniques based on the dataset sizes used for training. A comprehensive study is made on the positive aspects of various augmentation techniques using the sizes of the data. About 8 instances are taken into consideration with the size of the training set varying between 15K and 90K. Fig.3 shows that the proposed work performs positively in the low-data regime, resulting in positive improvement of augmentation in the images. It is also observed that as the number of data for training increases, there is a significant improvement observed using the proposed model when compared with other methodologies that do not use adversarial augmentation.

5. Conclusion

There are much research involved in recognition of handwriting in low-resource scripts. A number of work has also been carried out to provide solutions to the limitations imposed by generic data augmentation phenomenon. The proposed hybrid methodology will prove to be efficient in adding frameworks for word recognition and spotting, paving the way to deep networks which provide adequate information, even under low-data settings. Instead of using traditional methodology for augmenting handwriting text, the use of this methodology proves to be a reasonable solution which can be used to address differences in variation using datasets for training the inputs, thus proving a scalable solution. The amount of adaptability available when incorporating this methodology with respect to the opposing parameterization strategy serves to be useful in the implementation of rare unseen variation.

References

[1] Pham, V., Bluche, T., Kermorvant, C., & Louradour, J. (2014, September). Dropout improves recurrent neural networks for handwriting recognition. In 2014 14th international conference on frontiers in handwriting recognition (pp. 285-290). IEEE.

[2] Huang, K., Hussain, A., Wang, Q. F., & Zhang, R. (Eds.). (2019). Deep learning: fundamentals, theory and applications (Vol. 2). springer.

[3] Hamdan, Y. B. (2021). Construction of Statistical SVM based Recognition Model for Handwritten Character Recognition. Journal of Information Technology, 3(02), 92-107.

[4] Shibly, Mir Moynuddin Ahmed, Tahmina Akter Tisha, and Shamim H. Ripon. "Stacked Generalization Ensemble Method to Classify Bangla Handwritten Character." In Proceedings of International Conference on Sustainable Expert Systems, pp. 621-638. Springer, Singapore, 2021.

[5] Baldominos, A., Saez, Y., & Isasi, P. (2019). Hybridizing evolutionary computation and deep neural networks: an approach to handwriting recognition using committees and transfer learning. Complexity, 2019.

[6] Dutta, K., Krishnan, P., Mathew, M., & Jawahar, C. V. (2018, August). Improving cnn-rnn hybrid networks for handwriting recognition. In 2018 16th international conference on frontiers in handwriting recognition (ICFHR) (pp. 80-85). IEEE.

[7] Sinha, Gita, and Shailja Sharma. "Offline Handwritten Devanagari Character Identification." In International conference on Computer Networks, Big data and IoT, pp. 457-464. Springer, Cham, 2019.

[8] Altwaijry, N., & Al-Turaiki, I. (2021). Arabic handwriting recognition system using convolutional neural network. Neural Computing and Applications, 33(7), 2249-2261.

[9] Raj, Jennifer S., and Mr C. Vijesh Joe. "Wi-Fi Network Profiling and QoS Assessment for Real Time Video Streaming." IRO Journal on Sustainable Wireless Systems 3, no. 1 (2021): 21-30.

[10] Amin, Sujit S., and Lata Ragha. "Alphanumeric Character Recognition on Tiny Dataset." In International conference on Computer Networks, Big data and IoT, pp. 657-667. Springer, Cham, 2019.

[11] Suma, V. (2019). Computer vision for human-machine interaction-review. Journal of trends in Computer Science and Smart technology (TCSST), 1(02), 131-139.

[12] Carbune, V., Gonnet, P., Deselaers, T., Rowley, H. A., Daryin, A., Calvo, M., ... & Gervais, P. (2020). Fast multi-language LSTM-based online handwriting recognition. International Journal on Document Analysis and Recognition (IJDAR), 23(2), 89-102.

[13] Jacob, I. J. (2019). Capsule network based biometric recognition system. Journal of Artificial Intelligence, 1(02), 83-94.

[14] Boufenar, C., Kerboua, A., & Batouche, M. (2018). Investigation on deep learning for off-line handwritten Arabic character recognition. Cognitive Systems Research, 50, 180-195.

[15] Dhaya, R. "Light Weight CNN based Robust Image Watermarking Scheme for Security." Journal of Information Technology and Digital World 3, no. 2 (2021): 118-132.

[16] Koresh, H. James Deva, and Shanty Chacko. "Hybrid speckle reduction filter for corneal OCT images." In *International Conference on Image Processing and Capsule Networks*, pp. 87-99. Springer, Cham, 2020.

[17] Baldominos, A., Saez, Y., & Isasi, P. (2018). Evolutionary convolutional neural networks: An application to handwriting recognition. Neurocomputing, 283, 38-52.

[18] Smys, S., Chen, J. I. Z., & Shakya, S. (2020). Survey on Neural Network Architectures with Deep Learning. Journal of Soft Computing Paradigm (JSCP), 2(03), 186-194.

[19] Ahlawat, S., Choudhary, A., Nayyar, A., Singh, S., & Yoon, B. (2020). Improved handwritten digit recognition using convolutional neural networks (CNN). Sensors, 20(12), 3344.

[20] Jacob, I. J., & Darney, P. E. (2021). Design of Deep Learning Algorithm for IoT Application by Image based Recognition. Journal of ISMAC, 3(03), 276-290.

[21] Indumathi, V., and S. Prabakeran. "A Comparative Analysis on Sensor-Based Human Activity Recognition Using Various Deep Learning Techniques." In Computer Networks, Big Data and IoT, pp. 919-938. Springer, Singapore, 2021.

[22] Haoxiang, W., & Smys, S. (2020). MC-SVM Based Workflow Preparation in Cloud with Named Entity Identification. Journal of Soft Computing Paradigm (JSCP), 2(02), 130-139.

[23] Sueiras, J., Ruiz, V., Sanchez, A., & Velez, J. F. (2018). Offline continuous handwriting recognition using sequence to sequence neural networks. Neurocomputing, 289, 119-128.

[24] Manoharan, J. S. (2021). Capsule Network Algorithm for Performance Optimization of Text Classification. Journal of Soft Computing Paradigm (JSCP), 3(01), 1-9.

[25] Maalej, R., & Kherallah, M. (2020). Improving the DBLSTM for on-line Arabic handwriting recognition. Multimedia Tools and Applications, 79(25), 17969-17990.

[26] Sathesh, A., & Adam, E. E. B. (2021). Hybrid Parallel Image Processing Algorithm for Binary Images with Image Thinning Technique. Journal of Artificial Intelligence, 3(03), 243-258.

[27] Sen, S., Shaoo, D., Paul, S., Sarkar, R., & Roy, K. (2018). Online Handwritten Bangla Character Recognition Using CNN: A Deep Learning Approach. In Intelligent Engineering Informatics (pp. 413-420). Springer, Singapore.

Author's biography

Yasir Babiker Hamdan is presently working in International University of Africa (IUA), Khartoum, Sudan. His research is mainly focused on computer graphics, image processing, data mining, neural network algorithms and blockchain technologies.

A. Sathesh completed his master's degree in the year 2006 and has published several papers in national and international journals. His areas of interest include wavelets and multi-resolution transforms for image denoising. Currently, he is occupying an academic position in Eritrea after having worked in a reputed University in South India for the past 5 years. He is pursuing his research work in the area of complex wavelets for image approximations with a deep learning approach.