

Improving CNN-BGRU Hybrid Network for Arabic Handwritten Text Recognition

Sofiene Haboubi^{1,*}, Tawfik Guesmi², Badr M Alshammari², Khalid Alqunun², Ahmed S Alshammari², Haitham Alsaif² and Hamid Amiri¹

¹Signals Images and Information Technologies Lab. University of Tunis El Manar, National Engineering School of Tunis, BP 37 Le Belvedere, Tunis, 1002, Tunisia

²Department of Electrical Engineering, College of Engineering, University of Ha'il, Ha'il, 2440, Saudi Arabia

*Corresponding Author: Sofiene Haboubi. Email: sofiene.haboubi@enit.utm.tn

Received: 27 February 2022; Accepted: 06 May 2022

Abstract: Handwriting recognition is a challenge that interests many researchers around the world. As an exception, handwritten Arabic script has many objectives that remain to be overcome, given its complex form, their number of forms which exceeds 100 and its cursive nature. Over the past few years, good results have been obtained, but with a high cost of memory and execution time. In this paper we propose to improve the capacity of bidirectional gated recurrent unit (BGRU) to recognize Arabic text. The advantages of using BGRUs is the execution time compared to other methods that can have a high success rate but expensive in terms of time and memory. To test the recognition capacity of BGRU, the proposed architecture is composed by 6 convolutional neural network (CNN) blocks for feature extraction and 1 BGRU + 2 dense layers for learning and test. The experiment is carried out on the entire database of institut für nachrichtentechnik/ecole nationale d'ingénieurs de Tunis (IFN/ENIT) without any preprocessing or data selection. The obtained results show the ability of BGRUs to recognize handwritten Arabic script.

Keywords: Arabic handwritten script; handwritten text recognition; deep learning; IFN/ENIT; bidirectional GRU; neural network

1 Introduction

Optical Character Recognition (OCR) is the process of converting text image documents into a digital form easy for manipulation, indexing, preserving and searching. Early OCR systems are based on customized systems relying on the verification of each character image. Due to major changes within writing style, their accuracy is very low. Moreover, their training process is very slow. First OCR systems are traced back to 1914, where a system based on reading devices for blinds was developed [1,2].



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

OCR systems are widely used within multiple applications, basically data entry for business documents such as checks, passports and archives storing [3]. They are also used within automatic number recognition within traffic systems and historical documents storing. Moreover, they are widely used within controlling automatic systems.

OCR within Latin writing styles occupies the most research development within the field. Multiple approaches were established within this scope. Through the past decades within the development of Latin OCR engines, different approaches were used, including the development of handcrafted attributes, the use of hidden Markov models and neural networks. In recent years, deep neural networks are the most used techniques in the field due to their high accuracy within classification tasks [1,2].

Character recognition, in effect, means decoding any printed or handwritten text, or symbolic information. From the writing signal in its various forms (printed or handwritten, online or offline) [4], to decision making by a system, there are a few steps to be implemented: preprocessing, segmentation, feature extraction, learning and recognition. While the preprocessing and segmentation steps are not mandatory for character recognition systems.

The Arabic language is one of the most popular and widely used languages in the world. Standard Arabic script is a strictly semi-cursive script that consists of 28 base letters, 12 additional special letters, and eight diacritics, written from right to left.

In both printed and handwritten form, the Arabic writing is a semi-cursive. Characters of the same string (or PAW: part of an Arabic word) are ligated horizontally and vertically, which obstructs any attempt at segmentation into characters. The shape of a Arabic character change her form depending on its position in PAW. In addition, more than half of Arabic characters contain diacritics (1, 2 points or 3 points) in their form. These points can be below or above the character, but they can never be both at the same time. Several characters can be of the same size, but different numbers and/or positions of diacritics. Most letters can change their shape depending on their location at the word level (start, middle, end, or isolated), where each character can have up to four different shapes that increase the number of patterns (Fig. 1).

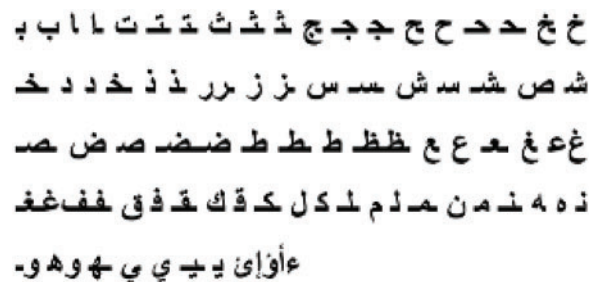


Figure 1: Different shapes of Arabic printed characters

Arabic character recognition is a research field interested by many works. Several approaches are proposed to recognize text from image input. The important approaches are presented in the competition international conference on document analysis and recognition (ICDAR/IFN-ENIT) (database for handwritten Arabic words) [5]. The recent works show the use of new architecture of text recognition can be a solution to improve best recognition accuracy that actual.

CNN were used by several researchers in vision computing, signal processing, pattern recognition and other fields. The obtained results are achieved with important success for some problem and

promising for other. Many contributions use the CNN structure for feature extraction and classification. The combination of CNN with other approaches is applied into many works, text recognition among them. Recent approaches that have shown good results are especially bidirectional long short-term memory–connectionist temporal classification (BLSTM-CTC) and multidirectional long short-term memory (MDLSTM). Their applications on Arabic texts give recognition rates respectively 92.43% [6] and 89.9% [7] on the IFN/ENIT database.

Currently, the use of gated recurrent unit (GRU) in pattern recognition can give good results. In recognition of Arabic characters, it is recently used by Mohd [8]. The proposed model uses CNN for the extraction of the features and implements BLSTM and BGRU for the recognition phase. The results obtained on Quranic printed text are promising with an accuracy 98%.

Recently, the GRUs and especially the BGRUs, have succeeded in solving several problems, such as, recognition of people from electrocardiogram (ECG) signals [9] and Arabic named entity recognition [10]. BGRU are also used in the field of text recognition, such as, Russian handwritten text recognition [11], but in handwritten Arabic text with large vocabulary, not yet done. Therefore, this work aims to improve handwritten Arabic text recognition accuracy by using the bidirectional GRU where two contributions can be noticed. We propose in the first main contribution is representing the IFN/ENIT database as big data with high complexity of recognition in handwritten Arabic text. The second main contribution is to improve the efficacy of the BGRU in hybrid network with CNN to recognize handwritten cursive text.

The remainder of this paper is organized as follows. A research background is investigated in Section 2, where we present a quick review on similar methods and bidirectional GRU techniques. The proposed optical model is developed in Section 3, where is based on CNN and BGRU. Materials and methods are presented in Section 4, by exploring the different sets of the IFN/ENIT database and presenting the used environment of experiments. Finally, main conclusions are presented in Section 5.

2 Background

2.1 Literature Review

The need for databases is for learning the developed system, as well, to test and diagnose the results found. Several databases are created to achieve this goal, but each has its own specificity. Depends on the problem to be solved, we choose the database. Among these databases, we are interest to handwritten Arabic script: IFN/ENIT [12], agriculture and horticulture development board (AHDB) [13], Urdu printed text images (UPTI) [14], king Fahd university of petroleum and minerals handwritten Arabic text (KHATT) [15], Alif [16], AHDB/Ftr [17], . . .

Due to the complexity of Arabic OCR systems, each approach addresses specific task. These works may be divided into three approaches. The first-class addresses font challenges in general such as binarization and image enhancement tasks. The second class deals with presented challenges by cursive scripts. Therefore, they address clean databases and assume the presence of well-defined separators. The third type deals with small word classes; and opt to segment and classify input data simultaneously. Despite the major differences between Latin and Arabic styles, the state-of-the-art architectures share the same standard processing style. These major differences impose multiple changes within the global processing flow of OCR systems, to be able to treat Arabic handwritten documents.

Among the handwriting recognition systems, we can group them into three types of systems:

- Hidden Markov models (HMM) which have shown a capacity for learning on structural sequences.
- Hybrid neuro-Markov models, which have made it possible to better model the local and global character of writing.
- Long short-term memory (LSTM) architectures, which even better integrate this ability to mix local and global, to optimize a decision over a complete sequence.

Average accuracies exceeded 99%, is a result obtained by Rahal [18]. This evaluation is tested on three different datasets (KHATT, mixed national institute of standards and technology (MNIST) and Arabic printed text image (APTI)). Her system for text recognition is based on statistical features by using bag of features model (BoF) and sparse auto-encoder (SAE). The HMM is used for recognition step.

Using OCR techniques, Zhang [19] developed a CAPTCHA recognition system with 98.08% accuracy rate. The proposed system uses the deep convolutional neural network (DCNN) and the convolutional recurrent neural network (CRNN). Also, CNNs for feature extraction and GRUs for sequencing are used by Suvarnam [20] for character recognition without going through the segmentation step. On a database of digits, Suvarnam showed the capacity of its system by a recognition rate of 100%. By using CNN and extreme learning machine (ELM), Ali [21] has developed a system for recognizing handwritten numbers. The tests of this system are made on the MNIST digits database, which gave an accuracy between 99.6% and 99.8%

The combination of CNN and BLSTM is used on the architecture system proposed by Youssef [16,22]. Zayene [23] uses the MDLSTM network to recognize words from Arabic text in video dataset (AcTiv) with a performance rate 96.5%.

2.2 Convolutional Neural Network

A convolutional neural network is a neural network that uses linear operations, products of convolution. Each convolutional neural network contains at least one convolutional layer. In the discrete domain the convolution is represented by the equation (formula (1)):

$$s(x) = (f \circ g)(x) = \sum_a f(t) g(x - a) \quad (1)$$

The most used convolution is a 2D convolution. In this case, for an image input I and for a kernel K , the discrete convolution is written (formula (2)):

$$S(i, j) = (K \circ I)(i, j) = \sum_m \sum_n I(i - m, j - n) K(m, n) \quad (2)$$

Yann LeCun [24] introduced convolutional layers at the start of a neural network (NN) to extract the features of images in a relevant way through convolution kernels. Several advanced and recent CNN architectures are proposed by researchers, such as residual neural network (ResNet) [25], inception [26], dense convolutional network (DenseNet) [27], mobile networks (mobileNetv2) [28] . . . , or personalized architecture such as [29].

In supervised learning, data is presented to the input of the neural network which produces outputs. The output depends on the parameters linked to the architecture of the neural network: connectivity between layers, aggregation, and activation functions. The backpropagation is to correct error between the system outputs and the desired outputs.

Updating the parameter values of an NN is done via the error gradient learning algorithm called backpropagation. Each parameter of an NN is updated during the backpropagation of the error gradient. It therefore ends up converging to zero: there is then loss of gradient and it is no longer possible to update the parameters of the recurrent neural network (RNN) (the vanishing gradient problem). To solve this problem, it is necessary to use a more complex NN architectures such as LSTM or GRU.

2.3 Bidirectional GRU

LSTM neural networks limit the vanishing gradient problem by a set of 3 gates (forget gate layer, input gate layer and output gate). GRUs are like LSTMs but they only have two gates (reset gate and update gate) [30]. The reset gate determines how much information to forget and update gate defines how much information to keep.

In Fig. 2, the mechanism of GRU is presented where x_t represent the input vector at time t . The previous hidden state is presented by h_{t-1} and the output of network is presented by h_t , which is expressed by the following formulas (3)–(6) [30]:

$$r_t = \sigma (W_r \cdot [h_{t-1}, x_t]) \tag{3}$$

$$z_t = \sigma (W_z \cdot [h_{t-1}, x_t]) \tag{4}$$

$$\tilde{h}_t = \varphi (W_{\tilde{h}} \cdot [r_t * h_{t-1}, x_t]) \tag{5}$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \tag{6}$$

where r_t and z_t are respectively the update gate and the reset gate, and the sigmoid function is presented by σ . The weight is denoted by W , and the elementwise production is denoted by \cdot .

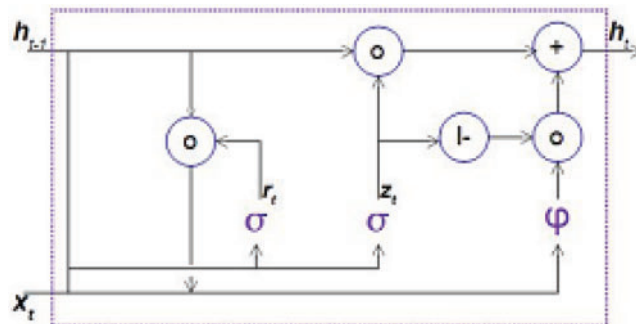


Figure 2: Structure of gated recurrent network

The BGRUs offer a mechanism to maximize the information processed [30,31]. It consists of concatenating the forward hidden layer and my backward hidden layer, which gives an output H_t (formula (7)).

$$H_t = \vec{h}_t + \overleftarrow{h}_t \tag{7}$$

3 Proposed Optical Model

Our approaches have two blocks: the convolutional block is to extract features from images and the recurrent block is to classify images (Fig. 3). The input images of the IFN/ENIT database are binary images containing names of Tunisian towns. From these images, the first block tries to extract the most relevant features by using CNNs layers. The convolutional block consists of 6 mini-blocks to design feature map with filters 16 and 32 for the first's layers and 48 for others. The applied kernel has a dimension 3×3 except the third and the fourth layer with kernel respectively 2×4 and 4×2 . After each convolutional layer we add rectified linear unit (ReLU) to rectify the small output values of the previous CNN layer. Most important values of the output matrix can be used in the next step. The last step of a mini-block is to calculate the largest values of each feature vector by renormalization (MaxPooling), the output represent the sampled vector of the most present features.

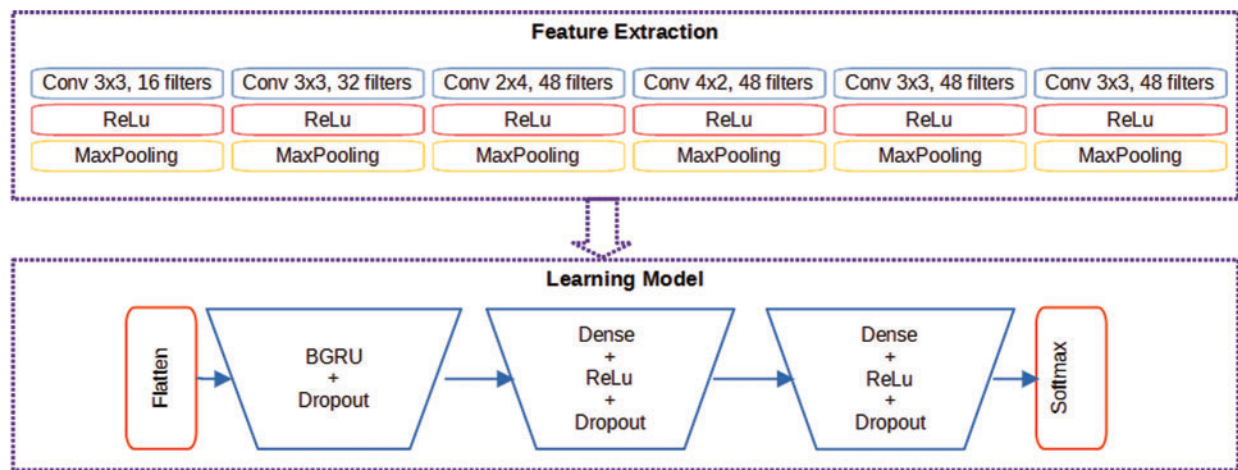


Figure 3: Proposed optical character recognition architecture

The second block contains 1 BGRU and 2 dense layers. The BGRU layer is followed by a dropout system and the dense layers are followed by ReLU and dropout. The dropout system is to ignore randomly some values to uniform with the input of the next step. Finally, to mapping the input image with the output decision, we need to add a layer with size equal to the number of town names.

4 Materials and Methods

4.1 IFNIENIT Database

IFN/ENIT is a database for handwritten Arabic words. It is used to develop Arabic handwritten words recognition systems. This database contains handwritten names of Tunisian cities written in Arabic by many writers (411 writers for only the first set). IFN/ENIT database contain in total 7 sets, about 7 thousand images for each. In the Fig. 4, we show few samples images of both datasets. In this work we use the sets a-b-c-d for training and the set e for testing. The details of sets are presented in the Tab. 1.

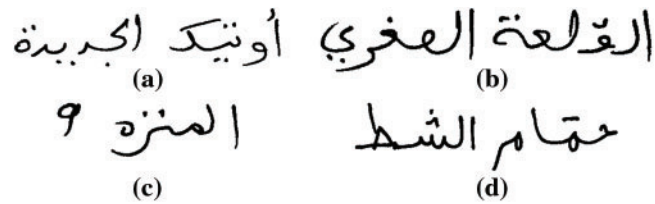


Figure 4: Examples of handwritten Arabic word from IFN/ENIT database

Table 1: Composition of IFN/ENIT database

	Set a	Set b	Set c	Set d	Set e	Sets abcde
Nb classes	823	835	829	858	733	937
Nb words	6537	6710	6477	6735	6033	32492
Nb PAWs	28298	29220	28391	29511	22640	138060
Nb characters	51984	53862	52155	54166	45169	257336

The sets a-b-c-d-e contains 937 Tunisian town names (937 classes). A town name can be composed by one Arabic word or more. Arabic digits can be present in the town name. The sets a-b-c-d-e contain 32492 Arabic words. The Arabic word can be composed by several connected characters (PAW), and a PAW can be composed by one character or more. Our datasets contain 257336 characters presented in 138060 PAWs. More details of each set are shown in the [Tab. 1](#) for example, in the [Fig. 4](#) contain four Tunisian town names writing with Arabic handwriting script. from right to left, the first name ([Fig. 4a](#)) contains two words. The first word contains three PAWs with one character for the PAW1, one character for the PAW2 and three characters for the PAW3.

The [Fig. 5](#) shows the number of samples for each town name that presented by her ZIP code. The ZIP code is used as a class name. Most classes have samples counts less than 50, while others sometimes exceed 300 samples. The number of samples is an important criterion in recognition, so with the imbalanced data we can analyze her impact on recognition system.

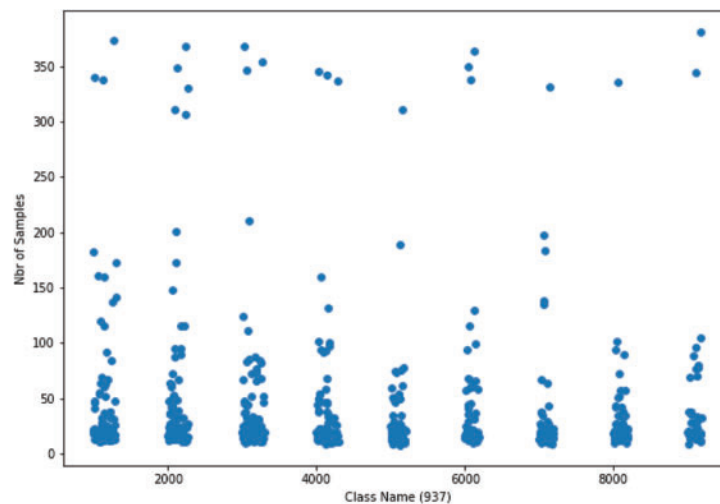


Figure 5: IFN/ENIT sets: a-b-c-d-e: number of samples by class name

The Fig. 6 shows the varieties of number of PAWs by image. More than 1000 images contain one PAW and more than 8000 images contain 4 PAWs. In total, the datasets have 138060 PAWs to be recognized varies between 1 and 10 PAWs for each image. This variety of number of PAWs increases the problem of recognition to build an efficient recognition system. This variety can be seeing also in the number of characters by sample. The Fig. 7 shows the number of characters by image. The little name size is composed by 3 characters and the longest by 17 characters. The most samples are composed by a character count between 4 and 10.

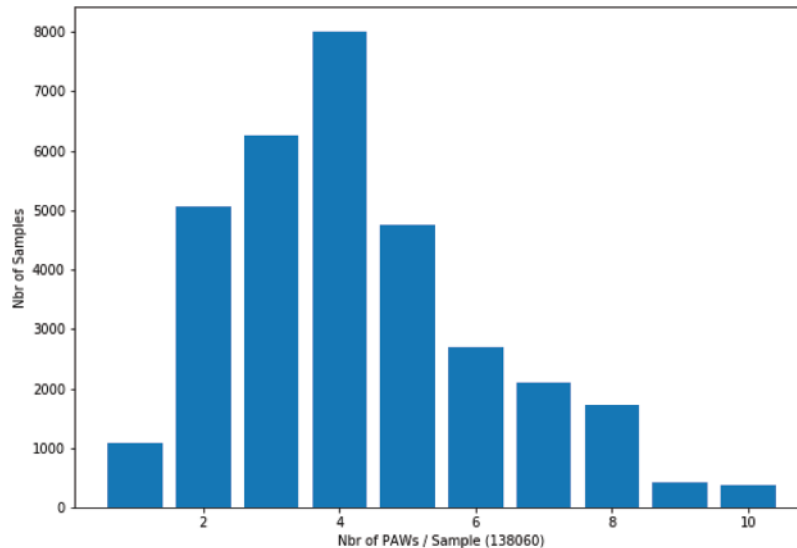


Figure 6: IFN/ENIT sets: a-b-c-d-e: number of samples by image weight (in PAWs)

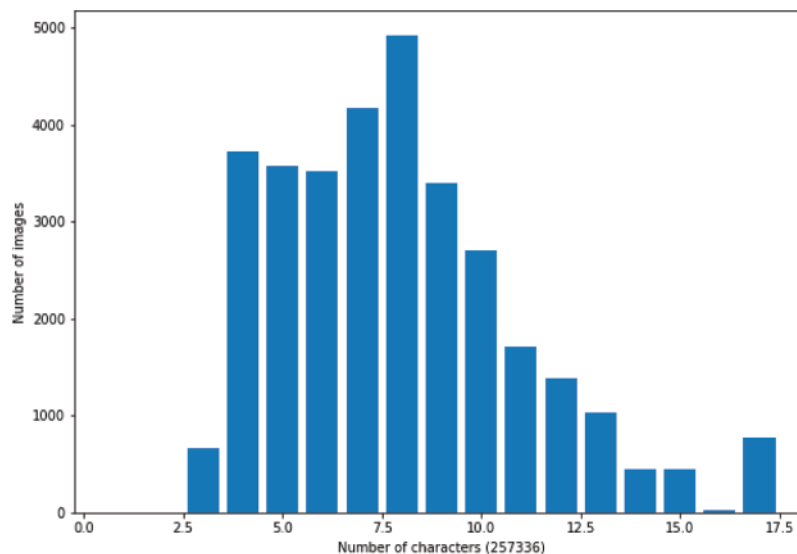


Figure 7: IFN/ENIT sets a-b-c-d-e: number of samples by image weight (in characters)

4.2 Experimental Evaluation

The test of our model was done on the IFN/ENIT database with its sets a, b, c, d and e. Each set contains mainly two folders: truth folder and handwritten text binary images folder. The first folder contains truth files, each file corresponds to an image from the second folder. The truth file contains information on the image (Tab. 2), the written text, ZIP code, writer's code, ... We initially started with processing truth files, extracting useful information (such as ZIP code). We then created a table mainly containing the image and its ZIP code. The ZIP code will be used as a class to recognize. From IFN/ENIT, we used sets a, b, c, and d (26459 images) for training and set e (6033 images) for test which is 19% of all data.

Table 2: Example of truth file of IFN/ENIT database

Label	Description
COM:	IFN/ENIT-database truth (label) file
COM:	http://www.ifnenit.com
COM:	IfN, TU-BS
COM:	bi43_047.tif coming from pb265_1.tif
X_Y	782 113
BDR:	begin data record
LBL:	ZIP:1273;AW1:Óíî.ÀÈÑÇáíã.ÇáÒãøÇÑ;AW2:seB yaM daE yaA aaA baB raE aaA heB yaM maE aaA laB zaE heB aaE raA ;QUA:YB1;ADD:P10
CHA:	17
BLN:	61, 57
EDR:	end of data record

The images received by our model are raw binary images, no preprocessing is applied apart from the resizing of the image to 256×56 pixels. The training and testing phase are implemented on a machine Intel Xeon E5-3.6 GHz with Windows operating system which is equipped with 32GB memory and GPU Nvidia Quadro M4000 8GB. Accuracy is used as the main success factor for our proposed model.

By comparing by the results in the literature; we have selected those which are tested on the IFN/ENIT database, and which used similar architectures. Among these architectures, the Tab. 3 shows that the best result is for the use of CNNs for the recognition of Arabic writing with a recognition rate of 97.07%. Considering the consumption in memory and time, other architectures are proposed in hybrid with the CNN such as BLSTM + CNN which gave a recognition rate of 92.21%. The recognition rates in the bibliography in the majority do not use the entire IFN/ENIT database, but rather a part of the database. The selection of the part used in the training reduces the number of classes used, also allows to control the number of samples of each class. The selection of a part of the database for training and testing complicates the comparison between the proposed models. Our model is tested on the entire database, which gives a recognition rate equal to 86.78%, but by selecting classes to be recognized according to her number of samples can attempt 100%. According to the evaluation results, we noticed that the number of samples in a class is very important for learning. Most of the false recognition are in the class which presents <10 samples.

Table 3: Methods based on CNN architectures

Reference	Design model	Recognition rate
[32]	CNN-BLSTM + CTC	92.21%
[33]	CNN	97.07%
[34]	multi-column deep neural network (MCDNN)	91.50%
[35]	deep belief network (DBN)	94.99%
[36]	CNN-DBN	95.20%
[37]	AlexNet	95.60%
Our	BGRU-CNN	86.78%–100%

Number of samples in the learning phase is a very interesting factor for the success of an intelligent system. This is also proved by our proposed system. [Tab. 4](#) shows the importance of number of samples in learning step and the [Tab. 5](#) shows the importance of number of samples per class to recognize a word. For example, the town name “المنزه” which corresponds to ZIP Code 1004, has 43 samples for training from sets a-b-c-d and 5 samples for test from set e. The Arabic town name “المنزه” has a recognition rate = 100% (correct classified: 5 and incorrect classified: 0).

Table 4: Recognition rate according to the variation between the number of samples in training and test steps. The samples of train set and the set has been selected randomly from the hole IFN/ENIT database

Experiment	Train-test	Nb train	Nb test	Recognition rate
1	50–50	16246	16246	50.14%
2	60–40	19495	12997	62.76%
3	70–30	22744	9748	83.28%
4	80–20	25993	6499	89.44%
5	90–10	29242	3249	96.54%

Table 5: Impact of the number of samples on the recognition rate of our model

Experiment	Nb of samples by class	Nb of class	Nb train	Nb test	Recognition rate
1	> 30	199	19966	5	100.0%
2	> 20	252	21208	141	100.0%
3	> 10	375	22855	985	98.43%
4	> 7	586	24516	2790	92.61%
5	> 5	914	26369	5801	89.24%
6	all	937	26459	6033	86.78%

From [Tabs. 4 and 5](#), it can be concluded that the number of samples is an important factor for the recognition of the handwritten Arabic word. This is noticed if we order the town names according to their recognition rates. Most who have a higher number of samples in the learning phase are the ones who are the highest accuracy.

[Tab. 5](#) shows the impact of the number of samples by town name. What is remarkable, that the number of samples of the learning phase is insufficient. For this reason, researchers have chosen to process and select classes for learning, by eliminating words that do not have an insufficient number of samples. Certainly, this increases the accuracy of the recognition system.

5 Conclusion

Arabic script is written from right to left with 28 letters with + 100 shapes. The Arabic character shapes vary according to their position in the word; This becomes more complex if the writing is handwritten. Several handwritten Arabic script recognition systems have been offered for several years. Considering the complexity of this script, good recognition results have been obtained recently. These systems are focused on the recognition rate, but they consume a lot of time and memory. The objective of this paper is to test a method which can save time and memory while maintaining a good recognition rate. The GRUs and especially the bidirectional-GRUs have shown good results in several research activities. By this present work, we want to improve the ability of BGRUs to recognize handwritten Arabic script.

The advantages of using BGRUs is the execution time compared to other methods that can have a high success rate but expensive in terms of time and memory [38]. To test the recognition capacity of BGRU, we propose an architecture composed by 6 CNN blocks for feature extraction and (1 BGRU + 2 Dense) layers for learning and test. The experiment is carried out on the entire IFN/ENIT database without any preprocessing or data selection. The obtained results show the ability of BGRUs to recognize handwritten Arabic script.

Funding Statement: This research was funded by the Deanship of the Scientific Research of the University of Ha'il, Saudi Arabia (Project: RG-20075).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] C. Reul, U. Springmann, C. Wick and F. Puppe, "Improving OCR accuracy on early printed books by combining pretraining, voting and active learning," *Journal for Language Technology and Computational Linguistics*, vol. 33, no. 1, pp. 3–24, 2018.
- [2] U. Springmann, C. Reul, S. Dipper and J. Baiter, "Ground truth for training OCR engines on historical documents in German fraktur and early modern Latin," *Journal for Language Technology and Computational Linguistics*, vol. 33, no. 1, pp. 97–114, 2018.
- [3] I. O. Omar, S. Haboubi and F. Benzarti, "Toward a binarization framework resolving the maghrebian font database challenges," in *Proc. SSD*, Monastir, Tunisia, pp. 462–466, 2020.
- [4] H. Amiri, S. Haboubi and H. Nakkach, "CHAKEL-DB: Online database for handwriting diacritic arabic character," in *Proc. VISIGRAPP*, Valletta, Malta, pp. 743–750, 2020.
- [5] M. Kherallah, N. Tagougui, A. M. Alimi, H. El Abed and V. Margner, "Online arabic handwriting recognition competition," in *Proc. ICDAR*, Beijing, China, pp. 1454–1458, 2011.

- [6] A. Graves, "Offline Arabic handwriting recognition with multidimensional recurrent neural networks," in *Guide to OCR for Arabic Scripts*, 1st ed., vol. 1, Germany: Springer, pp. 297–313, 2012.
- [7] R. Maalej and M. Kherallah, "Maxout into MDLSTM for offline arabic handwriting recognition," in *Proc. ICONIP*, Sydney, NSW, Australia, pp. 534–545, 2019.
- [8] M. Mohd, F. Qamar, I. Al-Sheikh and R. Salah, "Quranic optical text recognition using deep learning models," *IEEE Access*, vol. 9, pp. 38318–38330, 2021.
- [9] H. M. Lynn, S. B. Pan and P. Kim, "A deep bidirectional GRU network model for biometric electrocardiogram classification based on recurrent neural networks," *IEEE Access*, vol. 7, pp. 145395–145405, 2019.
- [10] N. Alsaaran and M. Alrabiah, "Arabic named entity recognition: A BERT-BGRU approach," *Computers, Materials & Continua*, vol. 68, no. 1, pp. 471–485, 2021.
- [11] A. Abdallah, M. Hamada and D. Nurseitov, "Attention-based fully gated CNN-BGRU for Russian handwritten text," *Journal of Imaging*, vol. 6, no. 12, pp. 1–23, 2020.
- [12] M. Pechwitz, S. Snoussi Maddouri, V. Märgner, N. Ellouze and H. Amiri, "IFN/ENIT-database of handwritten arabic words," in *Proc. CIFED*, Hammamet, Tunisia, pp. 129–136, 2002.
- [13] S. Al-Ma'adeed, D. Elliman and C. A. Higgins, "A database for arabic handwritten text recognition research," in *Proc. IWFHR*, Ontario, Canada, pp. 485–489, 2002.
- [14] N. Sabbour and F. Shafait, "A segmentation free approach to Arabic and Urdu OCR," in *Proc. SPIE*, Porto, Portugal, 2013.
- [15] S. A. Mahmoud, I. Ahmad, W. G. Al-Khatib, M. Alshayeb, M. T. Parvez *et al.*, "KHATT: An open arabic offline handwritten text database," *Pattern Recognition*, vol. 47, no. 3, pp. 1096–1112, 2014.
- [16] S. Yousfi, S. A. Berrani and C. Garcia, "ALIF: A dataset for arabic embedded text recognition in TV broadcast," in *Proc. ICDAR*, Nancy, France, pp. 1221–1225, 2015.
- [17] J. Ramdan, K. Omar, M. Faidzul and A. Mady, "Arabic handwriting data base for text recognition," *Procedia Technology*, vol. 11, pp. 580–584, 2013.
- [18] N. Rahal, M. Tounsi and A. M. Alimi, "Auto-encoder-BoF/HMM system for arabic text recognition," *ArXiv*, vol. abs/1812.03680, 2018.
- [19] R. Zhang, X. J. Wu, L. T. Qiu and Z. C. Yang, "OCR with a convolutional neural networks integration model in machine vision," in *Proc. ICDIP*, Shanghai, China, 2018.
- [20] B. Suvarnam and V. Sarma, "Combination of CNN-GRU model to recognize characters of a license plate number without segmentation," in *Proc. ICACCS*, Coimbatore, India, pp. 317–322, 2019.
- [21] S. Ali, J. Q. Li, Y. Pei, M. S. Aslam, Z. Shaukat *et al.*, "An effective and improved CNN-ELM classifier for handwritten digits recognition and classification," *Symmetry-Basel*, vol. 12, no. 10, pp. 1–15, 2020.
- [22] S. Yousfi, S. A. Berrani and C. Garcia, "Deep learning and recurrent connectionist-based approaches for arabic text recognition in videos," in *Proc. ICDAR*, Nancy, France, pp. 1026–1030, 2015.
- [23] O. Zayene, S. M. Touj, J. Hennebert, R. Ingold and N. E. Ben Amara, "Multi-dimensional long short-term memory networks for artificial arabic text recognition in news video," *IET Computer Vision*, vol. 12, no. 5, pp. 710–719, 2018.
- [24] Y. Lecun, L. Bottou, Y. Bengio and P. Haffner, "Gradient-based learning applied to document recognition," *IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [25] K. M. He, X. Y. Zhang, S. Q. Ren and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, Las Vegas, NV, USA, pp. 770–778, 2016.
- [26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proc. CVPR*, Las Vegas, NV, USA, pp. 2818–2826, 2016.
- [27] G. Huang, Z. Liu and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. CVPR*, Honolulu, HI, USA, pp. 2261–2269, 2017.
- [28] X. R. Zhang, J. Zhou, W. Sun and S. K. Jha, "A lightweight CNN based on transfer learning for COVID-19 diagnosis," *Computers, Materials & Continua*, vol. 72, no. 1, pp. 1123–1137, 2022.
- [29] M. Krendzelak and F. Jakab, "Hierarchical text classification using CNNs with local approaches," *Computing and Informatics*, vol. 39, no. 5, pp. 907–924, 2021.

- [30] L. Q. Li, L. Yang and Y. Y. Zeng, "Improving sentiment classification of restaurant reviews with attention-based Bi-GRU neural network," *Symmetry-Basel*, vol. 13, no. 8, pp. 1–16, 2021.
- [31] Y. W. Lu, R. P. Yang, X. P. Jiang, D. Zhou, C. S. Yin *et al.*, "MRE: A military relation extraction model based on BiGRU and multi-head attention," *Symmetry-Basel*, vol. 13, no. 9, pp. 1–15, 2021.
- [32] A. Lawgali, M. Angelova and A. Bouridane, "A framework for arabic handwritten recognition based on segmentation," *International Journal of Hybrid Information Technology*, vol. 7, pp. 413–428, 2014.
- [33] A. Poznanski and L. Wolf, "CNN-N-gram for handwriting word recognition," in *Proc. CVPR*, Las Vegas, NV, USA, pp. 2305–2314, 2016.
- [34] R. Almodfer, S. W. Xiong, M. Mudhsh and P. F. Duan, "Multi-column deep neural network for offline arabic handwriting recognition," *Artificial Neural Networks and Machine Learning*, vol. 10614, pp. 260–267, 2017.
- [35] N. Essa, E. El-Daydamony and A. A. Mohamed, "Enhanced technique for arabic handwriting recognition using deep belief network and a morphological algorithm for solving ligature segmentation," *ETRI Journal*, vol. 40, no. 6, pp. 774–787, 2018.
- [36] A. Khemiri, A. K. Echi and M. Elloumi, "Bayesian versus convolutional networks for arabic handwriting recognition," *Arabian Journal for Science and Engineering*, vol. 44, no. 11, pp. 9301–9319, 2019.
- [37] T. M. Ghanim, M. I. Khalil and H. M. Abbas, "Comparative study on deep convolution neural networks DCNN-based offline arabic handwriting recognition," *IEEE Access*, vol. 8, pp. 95465–95482, 2020.
- [38] R. Cahuantzi, X. Chen and S. Güttel, "A comparison of LSTM and GRU networks for learning symbolic sequences," in *Proc. ICONIP*, Bali, Indonesia, pp. 1–12, 2021.