On the use of the convolutional autoencoder for Arabic writer identification using handwritten text fragments

Amina Briber[0000-0003-1785-6406] and Youcef Chibani[0000-0002-7957-7456]

[1] Laboratoire d'Ingénierie des Systèmes Intelligents et Communicants (LISIC), Faculty of Electronics and Computer Science, University of Science and Technology Houari Boumediene (USTHB), 32, El Alia, Bab Ezzouar, 16111, Algiers, Algeria

{abriber,ychibani}@usthb.dz

Abstract. Convolutional autoencoders (CAE) are designed to reconstruct the input image to the output in a near-perfect way via a compact data namely encoded data containing relevant features. The encoded data can be used in various applications as for compressing or classifying the image. The present paper tries to investigate the use of the CAE for writer identification using handwritten text fragments. Hence, the CAE is used for generating features, which is fed to the distance-based classifier. Experimental evaluation is performed on the well-known IFN/ENIT dataset containing 411 writers. During training, a subset is selected from the 411 writers containing only 11 writers allowing to produce a lite CAE. Experimental results show an identification rate of 92.70% using the whole dataset when the feature vector is appropriately normalized.

Keywords: Writer identification, Handwritten, Text fragment, Convolutional autoencoders.

1 Introduction

In recent decades, the writer recognition has been one of the most challenging and fascinating research areas in the field of individual recognition. The hypothesis that writing is an individualistic act has been proven by psychologists and psychoanalysts [1]. The writer recognition is divided in two categories: writer identification and writer verification. This study focuses on the writer identification problem, for which several studies have been reported using the whole document [2], paragraphs [3], lines of text [4], words [5], characters [6] and recently text fragments [7,10]. In certain applications as for instance in forensics, few data are often available and therefore, the design of the writer identification system based on text fragments is considered as an interesting alternative way. Lastly, the writer identification from the text fragment has reached very interesting reliability levels as claimed by several authors [7,8]. The writer identification on small text fragments is also effective in absence of the whole document.

Usually, a writer identification system is composed of various modules, which are preprocessing, feature generation, classification and decision. The feature generation is the cornerstone of the system. Indeed, this module aims to represent a handwritten document by a set of features to describe the writing style of the writer [7]. Hence, the feature generation can be performed via two approaches: handcrafted and feature learning methods. The handcrafted approach consists of manually developing targeted descriptors to extract only the desired information. Several descriptors have been developed and used in this context, as for instance, the texture namely LBP, LTP and LPQ [7]. In contrast, the feature learning approach uses deep learning algorithms to extract all the information. These algorithms are basically based on neural networks such as recurrent neural networks (RNN), convolutional neural networks (CNN) [11] or convolutional autoencoders (CAE) [12].

The present paper aims to explore for the first time the use of the convolutional autoencoder (CAE) to extract features using handwritten text fragments. In contrast to CNN, the CAE does not require many data for training, which is considered as an advantage in real applications where, in certain circumstances, few data are available. For the classification module, the distance-based classifier is the most used to address the writer identification studies [8], since it offers an open system and quick execution time [8]. In this context, various distance kinds are used for classification when performed on handwritten text fragments [7-10]. In this paper, the simple Euclidean distance is used for writer identification using handwritten text fragments.

This paper is then organized as follows: Section 2 presents a brief review of convolutional autoencoders. Section 3 describes the proposed system. Section 4 devotes the experimental results and comparative analysis from the state of the art on text fragments for the Arabic writer identification. Finally, Section 5 presents out the conclusion and suggestions for future works.

2 Brief review of the convolutional autoencoder

The standard autoencoder is an unsupervised neural network [13] designed to reconstruct the input data to the output in a near-perfect way via a compact data namely encoded data [14]. Formally, the autoencoder involves two parts, which are an encoding part and a decoding part as shown in Fig.1.

The encoder allows producing an encoding data containing relevant features that are used next for reconstructing the image. Several kinds of autoencoders are proposed in the literature including sparse autoencoders [15], denoising autoencoders [15], convolutional autoencoders, variational autoencoders and so on [16].

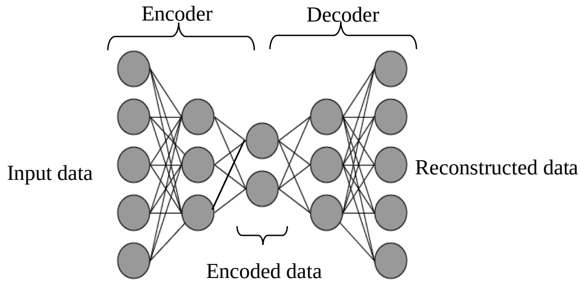

Fig. 1. Autoencoder architecture.

Convolutional autoencoders (CAE) are well adapted for image processing for extracting all features from the input image [17]. CAE are based on a standard autoencoder architecture having convolutional encoding and decoding layers [18]. In convolutional autoencoders, the encoding process involves to performing a convolution operation on the input layer image in order to extract local features, thus obtaining a hidden layer input. The decoding process performs a deconvolution operation on the hidden layer data, so that the output image is reconstructed to the size of the input image [19].

This paper tries to investigate the use of the convolutional autoencoder as a feature generator associated to distance-based classifier for writer identification based on handwritten text fragments.

3 Description of the proposed system

The proposed paper aims to propose an open system for writer identification using text fragments. Therefore, each writer is represented by a set of text fragments. As shown in Fig. 2, the proposed system contains two main modules: the feature generation module and the classification module.

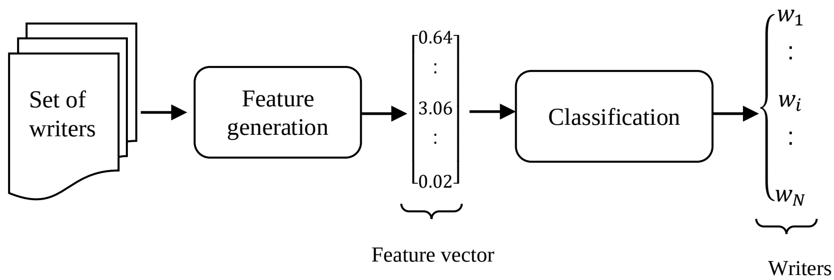

Fig. 2. Proposed writer identification system.

The feature generation is performed *via* the convolutional autoencoder, which is used as a feature generator for producing a feature vector. The resulting one is retrieved and then used as an input for the classification module using the Euclidean distance for identifying a writer. For better clarity, a more deeply description of each module is reported in the following sections.

3.1 Feature generation

The feature generation is considered as a crucial step in the writer identification system. Its role is to extract features contained into the text fragments in order to form a feature vector that will be used as input for the classification module. As shown in Fig. 3, the design of the feature generator is performed into two steps recovering the feature vector *via* the CAE and then followed by a normalization way.

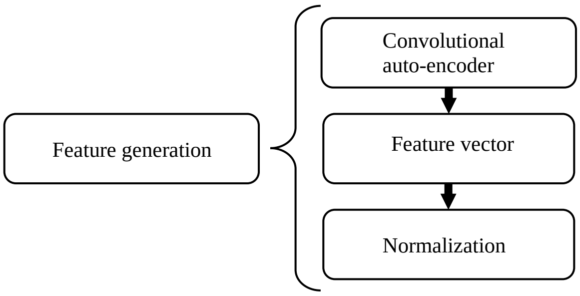

Fig. 3. Feature generation steps.

The CAE is used for representing the text fragment in a compact form, which defines the feature vector as shown in Fig. 4. The feature vector is considered representative when the output fragment is similar to the input fragment according to a measure criterion. The representative feature vector is found during the design step.

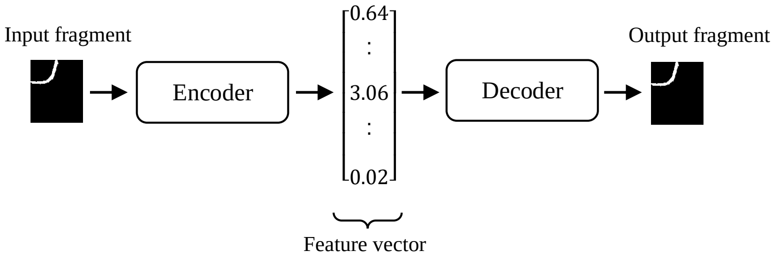

Fig. 4. Convolutional autoencoder for feature extraction.

The encoder and decoder of the proposed CAE are composed of convolution layers, an activation function (ReLU) and a pooling layer. Table 1 shows the architecture of the used CAE.

To ensure the homogeneity of the values contained into the feature vector, a normalization is performed for redistributing values in the range between zero and one. The mathematical formula for normalization is defined as follows [20]:

$$y_n = \frac{x_n^2}{\sum_{n=1}^{P} x_n^2} \quad (1)$$

P is the size of the feature vector, while x_n and y_n represent the no normalized and normalized feature vectors, respectively.

Table 1. Architecture of the proposed CAE.

	Layer type	Number of Filters	Kernel size	Padding	Size of the output
Encode	Input image	-	-	-	100 × 100 × 1
	Convolution	5	3 × 3	Same	100 × 100 × 5
	ReLU	-	-	-	100 × 100 × 5
	MaxPooling	-	2 × 2	Same	50 × 50 × 5
	Convolution	10	3 × 3	Same	50 × 50 × 10
	ReLU	-	-	-	50 × 50 × 10
	MaxPooling	-	2 × 2	Same	25 × 25 × 10
	Convolution	20	3 × 3	Same	25 × 25 × 20
	ReLU	-	-	-	25 × 25 × 20
	MaxPooling	-	2 × 2	Same	13 × 13 × 20
Code	Flatten	-	-	-	3380
Decode	Convolution	20	3 × 3	Same	13 × 13 × 20
	ReLU	-	-	-	13 × 13 × 20
	UpSampling	-	2 × 2	Same	26 × 26 × 20
	Convolution	10	3 × 3	Same	26 × 26 × 10
	ReLU	-	-	-	26 × 26 × 10
	UpSampling	-	2 × 2	Valid	52 × 52 × 10
	Convolution	5	3 × 3	Same	50 × 50 × 5
	ReLU	-	-	-	50 × 50 × 5
	UpSampling	-	2 × 2	Same	100 × 100 × 5
	Convolution	1	3 × 3	Same	100 × 100 × 1
	Output image	-	-	-	100 × 100 × 1

3.2 Classification

Inspired from [7], the proposed study also uses a dissimilarity measure between fragments of reference writer storage in the database and fragments of the query writer.

Let $r_j, j = 1,..,card(R)$ as the feature vector of the fragment belonging to the reference writer R and $card(R)$ is the number of fragments, and let $q_i, i = 1,..,card(Q)$ as the feature vector of the fragment belonging to the query writer Q and $card(Q)$ is the number of fragments. The dissimilarity measure namely Γ is calculated in order to compare between two writers, which is defined as follows:

$$\Gamma(Q,R) = \frac{1}{card(Q)} \sum_{i=1}^{card(Q)} \min_{r_j \in R}(d(q_i, r_j)) \tag{2}$$

Where $d(q_i, r_j)$ is the distance between two fragments (q_i, r_j), which can be computed *via* the Euclidean distance defined as follows:

$$d(q_i, r_j) = \sqrt{\sum_{n=1}^{P} (q_{ni} - r_{nj})^2} \tag{3}$$

P defines the size of the feature vector.

For identifying a writer, the dissimilarity measure is performed against all writers stored in the database. A decision rule is then performed for selecting the minimal dissimilarity measure among all calculated dissimilarity measures. Formally, the index of the writer namely W_{index} is obtained through the following equation:

$$W_{index}(Q) = \arg\min_{k=1,..,K}(\Gamma(Q, R_k)) \quad (4)$$

Where K is the number of writers stored in the database.

4 Experimental results

4.1 Dataset description

For evaluating the proposed writer identification system, the well-known IFN/ENIT dataset is used. It includes 2,200 documents with more than 26,000 names of Tunisian cities and town villages written in Arabic collected from 411 different writers. In this paper, the text fragments are produced for training and testing according to Hannad et al [7] for a fair comparison. Fig. 5 shows some samples of the text fragments from the same writer. [21]

Fig. 5. Samples of the IFN/ENIT text fragments of the same writer.

4.2 Evaluation criterion

The performance of the system is evaluated using the identification rate (IR), which is equivalently to Accuracy. It is defined as the ratio of the number of well identified writers to the total number of writers as follows:

$$IR(\%) = \frac{\text{Number of writers well identified}}{\text{Total number of writers}} \times 100 \quad (5)$$

This measure is used for a fair comparison against the state-of-art.

4.3 Experimental design

For finding the optimal architecture of the CAE model offering the best identification rate, the first eleven (11) writers are selected for designing the proposed system while the remaining 400 writers are used for evaluation. The selected training fragments belonging to 11 writers are divided into two subsets. The first subset containing 80% of training fragments is used for training the CAE while the second subset containing 20% of training fragments is used for validation. Testing fragments of the eleven (11) writers are used for evaluating the robustness of the CAE. No data augmentation is performed during the training of the CAE. The number of writer's fragments varies from writer to another. Once the system is validated, the evaluation is performed on the whole dataset. For the training, the number of epochs is fixed at 350 and the binary cross entropy is used as a loss function. Fig. 6 shows the accuracy (Identification Rate) and the loss of the model for training and validation samples.

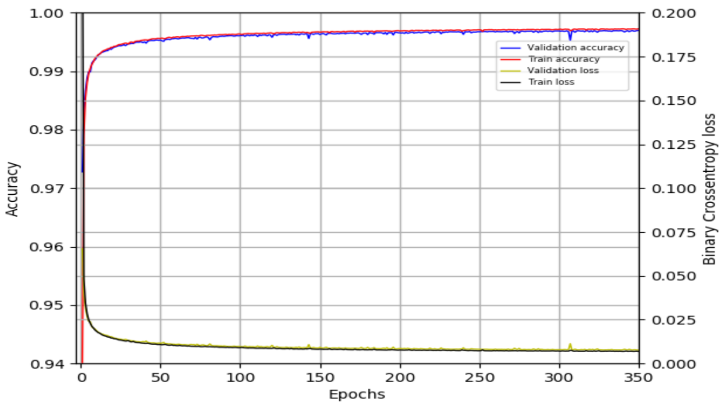

Fig. 6. Accuracy and the loss for training and validation.

As it can be seen, the model is well trained without overfitting and with a high accuracy, which means that the input image is reproduced on the output of the CAE as almost identical. Therefore, the feature vector retrieved from the CAE can be considered sufficiently relevant for using in the writer identification.

4.4 Experimental evaluation

The proposed system is evaluated on the whole IFN/ENIT dataset containing 411 writers. The model is evaluated firstly on a subset of the IFN/ENIT dataset containing 11 writers and the remaining writers are added progressively in order to keep an open system. At first, the evaluation is performed on the feature vector directly without normalization. Afterward, a normalization is performed on the feature vector as described in equation 2 in order to show its influence. Fig. 8 depicts the identification rate using normalized and no normalized feature vector, respectively.

As expected, the identification rate decreases when the number of writers increases. However, the normalized feature vector is better since it keeps its stability performance compared to the no normalized feature vector. It can be seen also that the normalized feature vector gives an encouraging identification rate of 92.70% with an increase of 6.57% from 86.13% using the feature vector without normalization. This experiment clearly shows that the normalization allows a considerable influence for improving the identification rate. Furthermore, the evaluation shows the stability performance of the proposed system against the number of writers. Indeed, when a new writer is added to the system, the identification rate remains almost stable.

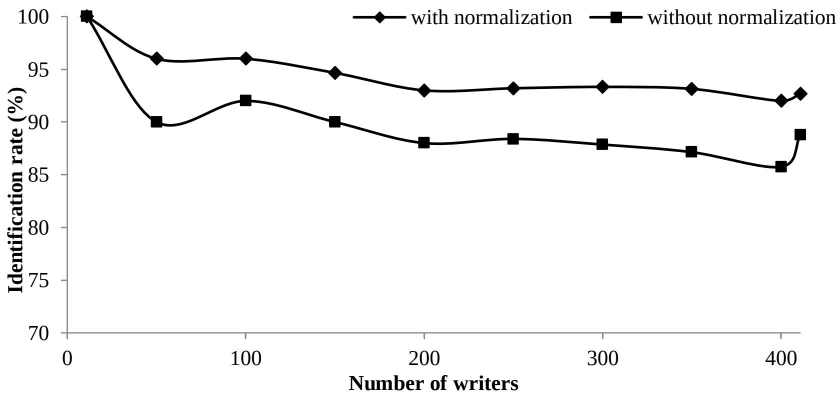

Fig. 8. Identification rate versus number of writers.

4.5 Comparative analysis

In order to situate the proposed work among the works carried out from the state-of-art, Table 2 shows only selected single systems (no combined systems) for a fair comparison when using the fragmented IFN/ENIT dataset. As can be seen, all existing single systems use handcrafted features whereas the proposed system use a feature learning method using the convolutional autoencoder (CAE).

Table 2. IR (%) obtained by the offered system against the state-of-the-art methods using text fragments for the IFN/ENIT dataset.

Method	**Reference**	**Feature Extraction**	**Classifier**	**IR (%)**
	Hannad et al [7]	LBP	Hamming distance	73.48
		LTP	Hamming distance	87.12
		LPQ	Hamming distance	94.89
	Hannad et al [9]	HOG	Hamming distance	86.62
	Hadjadji et al [8]	Run Length	k-Means	93.18
Handcrafted			k-Centers	88.07
			Hamming distance	92.45
	Hadjadji et al [8]	oBIF	k-Means	82.23
			k-Centers	78.34
			Hamming distance	90.51
	Hadjadji et al [8]	LPQ	k-Means	93.67
			k-Centers	88.32
Feature Learning	**Proposed**	**CAE**	**Euclidian Distance**	**92.70**

Hannad et al [7,9] used different descriptors associated to the Hamming distance classifier. The LPQ descriptor showed the best performance with an IR of 94.89%. While Hadjadji et al [8] evaluated several descriptors with different classifiers. The LPQ descriptor associated to the k-means classifier has giving an IR of 93.67%.

Through this comparison, it can be deduced that the IR obtained by the proposed system is ranked in 4th position, with an IR of 92.70%, compared to the rates of the existing systems.

It is worth noting that systems designed by Hannad et al [7] and Hadjadji et al [8] deleted a component from the LPQ descriptor to improve the identification rate. According to [7,8], the text fragments contain no-pertinent fragments that can influence the result. In the proposed system, all feature vector components are retained.

Among the existing systems, the proposed system is an open system. For instance, when a new writer is added to the system, the re-training of the CAE is not required. Furthermore, a small amount of data is required for training the CAE with only 11 writers without data augmentation.

The presented system is promising and offers an encouraging identification rate while it remains a very simple and lite system offering fast and effective results.

5 Conclusion

This paper proposed to investigate a new Arabic writer identification open system using convolutional autoencoder (CAE) and distance-based classifier from text fragments. The used approach is based on the CAE model for generating features performed on a subset of writers, each one is represented by all his fragments. In the sequel, the same model is used for identifying the query writer from all writers contained into the IFN/ENIT dataset using the distance-based classifier without retraining the model. The identification rate achieved 92.70% with a lite CAE model.

The use of CAE such as feature extractor for writer identification task-based text fragments shows an encouraging performance for capturing the relevant features of the writer style, despite the lite model used and the small amount of data used in the training step. For further works, an investigation is planned for reducing the size of the feature vector by trying to improve the performance.

Acknowledgement

This work was supported by the Direction Générale de la Recherche Scientifique et du Développement Technologique (DGRSDT) grant, attached to the Ministère de l'Enseignement Supérieur et de la Recherche Scientifique, Algeria.

References

1. Saks, M. J. Commentary on: Srihari SN, Cha SH, Arora H, Lee S. Individuality of handwriting. J Forensic Sci 2002; 47 (4): 856-72. Journal of forensic sciences, 48(4), 916-920 (2003).

2. Chawki, D., Labiba, S.-M. A texture based approach for Arabic writer identification and verification. In: Machine and Web Intelligence (ICMWI), 2010 International Conference on. IEEE, pp. 115–120. (2010).
3. Bulacu, M., Schomaker, L., Brink, A. Text-independent writer identification and verification on offline arabic handwriting. In Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) (Vol. 2, pp. 769-773). IEEE. (2007, September).
4. Abdi, M. N., Khemakhem, M. A model-based approach to offline text-independent Arabic writer identification and verification. Pattern Recognition, 48(5), 1890-1903. (2015).
5. Abdi M, Khemakhem M, Ben-Abdallah H. A Novel Approach for Off-Line Arabic Writer Identification Based on Stroke Feature Combination. 24th International. (2009)
6. Idicula, S. M. A survey on writer identification schemes. International Journal of Computer Applications, 26(2), 23-33. (2011).
7. Hannad, Y., Siddiqi, I., El Kettani, M. E. Y. Writer identification using texture descriptors of handwritten fragments. Expert Systems with Applications, 47, 14-22. (2016).
8. Hadjadji, B., Chibani, Y. Two combination stages of clustered one-class classifiers for writer identification from text fragments. Pattern Recognition, 82, 147-162. (2018).
9. Hannad, Y., Siddiqi, I., El Merabet, Y., El Youssfi El Kettani, M. Arabic writer identification system using the histogram of oriented gradients (HOG) of handwritten fragments. In Proceedings of the Mediterranean Conference on Pattern Recognition and Artificial Intelligence (pp. 98-102). (2016, November).
10. Tang, Y., Wu, X., Bu, W. Offline text-independent writer identification using stroke fragment and contour based features. In 2013 International Conference on Biometrics (ICB) (pp. 1-6). IEEE. (2013, June).
11. He, S., Schomaker, L. Fragnet: Writer identification using deep fragment networks. IEEE Transactions on Information Forensics and Security, 15, 3013-3022. (2020).
12. Zhu, Y., Wang, Y. An offline text-independent writer identification system with SAE feature extraction. In 2016 International Conference on Progress in Informatics and Computing (PIC) (pp. 432-436). IEEE. (2016, December).
13. Dong, C., Xue, T., Wang, C. The feature representation ability of variational autoencoder. In 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC) (pp. 680-684). IEEE. (2018, June).
14. Ng, A. Sparse autoencoder. CS294A Lecture notes, 72(2011), 1-19. (2011).
15. Bank, D., Koenigstein, N., Giryes, R. Autoencoders. arXiv preprint arXiv:2003.05991. (2020).
16. Guo, X., Liu, X., Zhu, E., Yin, J. Deep clustering with convolutional autoencoders. In International conference on neural information processing (pp. 373-382). Springer, Cham. (2017)
17. Gondara, L. Medical image denoising using convolutional denoising autoencoders. In 2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW) (pp. 241-246). IEEE. (2016, December).
18. Masci, J., Meier, U., Cireşan, D., Schmidhuber, J. Stacked convolutional auto-encoders for hierarchical feature extraction. In International conference on artificial neural networks (pp. 52-59). Springer, Berlin, Heidelberg. (2011, June).
19. Kundur, D., Hatzinakos, D. Blind image deconvolution. IEEE signal processing magazine, 13(3), 43-64. (1996).
20. Christlein, V. Handwriting analysis with focus on writer identification and writer retrieval. (2019).
21. Awaida, S. M., Mahmoud, S. A. State of the art in off-line writer identification of handwritten text and survey of writer identification of Arabic text. Educational Research and Reviews, 7(20), 445-463. (2012).