International Journal of Engineering

Journal Homepage: www.ije.ir

Holistic Persian Handwritten Word Recognition using Convolutional Neural Network

A. Zohrevand*, Z. Imani

Computer Engineering Department, Kosar University of Bojnord, Bojnord, Iran

PAPER INFO

Paper history:
Received 16 April 2021
Received in revised form 07 June 2021
Accepted 06 July 2021

Keywords:
Persian Handwritten Word Recognition
Convolutional Neural Network
End-to-end Learning Method
Transfer Learning
Persian Handwritten Dataset

ABSTRACT

Due to the cursive-ness and high variability of Persian script, the segmentation of handwritten words into sub-words is still a challenging task. These issues could be addressed in a holistic approach by sidestepping segmentation at the character level. In this paper, an end-to-end holistic method based on deep convolutional neural network is proposed to recognize off-line Persian handwritten words. The proposed model uses only five convolutional layers and two fully connected layers for classifying word images effectively, which can lead to a substantial reduction in the required parameters. The effect of various pooling strategies is also investigated in this paper. The primary goal of this article is to ignore handcrafted feature extraction and to attain a generalized and stable word recognition system. The presented model is assessed using two famous handwritten Persian word databases called Sadri and IRANSHAHR. The recognition accuracies were obtained at 98.6% and 94.6%, on Sadri and IRANSHAHR datasets respectively, and outperformed the state-of-the-art methods.

***doi**: 10.5829/ije.2021.34.08b.24*

1. INTRODUCTION

Handwriting recognition refers to the process of converting handwritten images into their corresponding editable files [1, 2]. Unlike printed texts, due to significant changes in writing style and shape, the skew or slant automatic handwritten text recognition is still a debatable subject in the pattern recognition and computer vision community [1]. Handwriting recognition has several applications such as bank cheque processing [3], the recognition of notes [4, 5], postal address recognition [6] and historical documents [7, 8] in various scripts including Indian, Chinese, Latin, Arabic, and Persian. Compared to Latin/Roman, Chinese and Japanese scripts in which texts are written separately, in Persian script, the texts are usually written cursively, which further complicates the recognition process [9].

Handwritten recognition methods are divided into two major groups: on-line and off-line methods [10]. While on-line recognition depends on pen movement coordinates and the pen trajectory of the letter, off-line recognition is based on the analysis of the text image [1]. Also, there are two methods for word recognition: segmentation-based and holistic methods [11]. In the segmentation-based, an input image is segmented into different sets of sub-words, and the word is recognized by its constituent units. In the holistic approach, however, an input image is recognized by its shape as a whole. Recently, Convolutional Neural Network (CNN) models have been widely used in various computer vision applications such as image segmentation, image classification, object detection and recognition due to their capability to directly extract high-level features from images [12]. In this article, attempts have been made to design a CNN model for holistic Persian off-line handwritten word recognition.

To the best of our knowledge, despite the excellent performance of CNN in a variety of computer vision applications, the use of CNN models for the holistic-based Persian off-line handwritten word recognition has received scant scholarly attention. Therefore, this study uses a novel CNN model to recognize Persian handwritten words. The main contributions of this article are as follows:

- Inspiring by [1], this paper proposed an end-to-end learning architecture that eliminates the need for

*Corresponding Author Institutional Email: a.zohrevand@kub.ac.ir
(A. Zohrevand)

Please cite this article as: A. Zohrevand, Z. Imani, Holistic Persian Handwritten Word Recognition using Convolutional Neural Network, International Journal of Engineering, Transactions B: Applications, Vol. 34, No. 05, (2021) 2028-2037

handcrafted feature extraction in holistic Persian off-line handwritten word recognition.
- Presenting a new Transfer Learning (TL) approach for holistic Persian off-line handwritten word recognition.
- Analysing the proposed method on two popular Persian handwritten word datasets called IRANSHAHR [13] and Sadri [14].
- Analysing errors on Sadri dataset for the first time.

The rest of this paper is formed as follows. Section 2 reviews the related works about Persian handwritten word recognition. The proposed methodology is explained in Section 3. Sections 4 presents details of the experimental results. Error analysis is described in Section 5. Discussion and comparison are presented in Section 6. Finally, conclusions and future works are explained in Section 7.

2. LITERATURE REVIEW

Several important studies have explored the handwritten word recognition in Persian script, which are reviewed in this section. Dehghan et al. [15] proposed a vector quantization method based on Self-Organizing Feature Map (SOFM) in order to recognize Persian handwritten words. In this work, the contour image of each word is scanned from right to left to extract features. SOFM was utilized to prepare a codebook and smoothen the distribution of observation probabilities. Then, a distinct Hidden Markov Model (HMM) was trained on each word. The same authors [16] proposed a similar approach based on fuzzy C-means. The database used in both studies was 198 classes from IRANSHAHR dataset. For the first and second methods, the recognition accuracies were 65% and 67.2%, respectively. Mozaffari et al. [17] proposed a lexicon reduction method based on single, double and triple dots. In their approach, the words with different dot patterns are discarded prior to the classification. This strategy not only enhances accuracy but also the accelerate recognition process. As an experimental result, a recognition accuracy of 73.61% was obtained for 200 classes of IRANSHAHR dataset.

Broumandnia et al. [18] proposed rotation and scale invariant features for the recognition of holistic Arabic/Persian handwritten words. In their approach, M-band wavelet transform was used for features extraction and the Mahalanobis distance for classification. The experimental results showed a 12% improvement in the recognition accuracy on the database provided by themselves. Arani et al. [10] combine the output of Left-to-Right (LtR) and Right-to-Left (RtL) HMMs to recognize Persian handwritten words. In their approach, the LtR and RtL drew on complement rules for decreasing errors in recognition accuracy. Imani et al. [2] used a sliding window that sweeps vertically across a word image for the extraction of intensity and directional gradient features. The features extracted from each window are then coded by the Self Organizing Map (SOM). Then, a distinct HMM was trained on each word. The recognition accuracy was obtained 69.07% on a database that contained 30,000 images from 300 formal words [19].

Arani et al. [11] extracted three feature groups including white-black transition, image gradient, and the chain code of contour from the input image. For each feature group, a discrete HMM was trained for each word. Finally, the outcomes of three HMMs are fused by a multilayer perceptron that is responsible for recognition classification. The recognition accuracy was obtained 89.06% on the 200 classes of IRANSHAHR dataset. Tavoli at al. [20] proposed new approach for extracting appropriate features to recognize Arabic/Persian handwritten words. In this method, the input image is divided into an $n \times m$ strip from which straight lines are extracted. Then, based on the location, number, angle, and straight line size, some geometrical features are extracted. The classification is conducted using the Support Vector Machine (SVM). Their proposed approach was evaluated on three databases: IBN SINA [21], IFN/ENIT [22] for Arabic and IRANSHAHR for Persian. Recognition accuracy of the proposed method was reported about 67.47, 86.22, and 80.78% for the IRANSHAHR, IBN-SINA, and IFN/ENIT dataset respectively . The above works are impressive, but since all of them employ manual methods for feature extraction, they are known as handcrafted features. The success of a recognition system primarily relies on the extraction of proper features from the word images. Generally, extracting handcrafted features are challenging, boring and in some case impossible for researchers. The end-to-end learning in the absence of handcrafted features is one of the remarkable characteristics of the CNN. Therefore, CNN-based models can be a good alternative for the recognition of Persian handwriting words.

LeCun et al. proposed the CNN architecture for the first time [23]. CNN models have been the subject of considerable attention in most computer vision applications [24-27]. These families of neural network, which fuse classification and feature extraction tasks, are intended to recognize images based on their scale, shift, and distortions. Safarzadeh et al. [28] proposed a novel approach according to the sequence labeling with CNN and Bidirectional Long Short Term Memory (LSTM). In this method, in order to ignore the segmentation step in segmentation-based methods, a Connectionist Temporal Classification (CTC) is used as the cost function. For feature extraction, the sequences of feature vectors are extracted by CNN model from an input word image. Then, the Recuurent Neural Network (RNN) model beside CTC cost function is utilized for input sequence

labeling as the classification task. According to the experimental results, the recognition accuracy was obtained 98.8% on the Sadri dataset, which contains 62,500 images from 125 word classes. Bonyani et al. [29] present ensemble method based on various CNN architectures for the recognition of handwritten Persian letters, digits and words. Specifically, an ensemble of various DenseNet [30] architectures and Xception [31] were utilized for word recognition. In experimental results, the recognition accuracy was reported 98.8% on Sadri dataset. These studies are summarized in Table 1.

3. PROPOSED METHODOLOGY

The general chart of the presented method is depicted in Figure 1. As shown in this figure, the presented approach consists of two major parts: preprocessing and training the CNN model. The preprocessing details and CNN model are explained in the following subsections.

3. 1. Preprocessing In this paper, two databases called Sadri and IRANSHAHR are used. There are word images of various sizes in the two datasets. Also, for word images, the region of interest (i.e. handwritten area) in each class is different. However, the proposed CNN model needs 224×224 images as the input size in the first layer. There are several methods for image resizing [14]. The simplest one is the resizing of all images to 224×224, but applying this method on all images with different sizes may deform structure of handwritten images. Sabzi at al. [32] proposed a simple approach for an efficient image scaling (resizing). In this approach, all images in each dataset are divided into two groups. In the first group, images with dimensions smaller than the standard size, are only padded with a background pixel. In the second group, for images with one or two dimensions greater than 224, ratio of the standard size to the bigger dimension is calculated, then the height and width of the input image are resized according to the ratio. The result of image resizing is shown in Figure 2. As shown in this figure, compared to the image resizing by normal scaling (middle column), this strategy (right column) does not change the structure of input images.

3. 2. Proposed CNN Architecture Dealing with images directly reduces the performance of existing neural networks. Thus, a handcraft feature extraction step

TABLE 1. Some seminal works on the recognition of Persian handwritten words

Ref.	Method	Dataset	Recognition method	Year
[15]	Self-Organization Map, discrete HMM	IRANSHAHR	Holistic	2001
[16]	Fuzzy vector quantization, discrete HMM	IRANSHAHR	Holistic	2001
[17]	Lexicon reduction, discrete HMM	IRANSHAHR	Segmentation-based	2008
[18]	M-band wavelet transform	Private	Holistic	2008
[10]	Fusion of Right-to-Left, Left-to-Right HMMs	IRANSHAHR	Holistic	2018
[2]	Image gradient, discrete HMM	Private	Holistic	2014
[11]	Discrete HMM, classifier fusion	IRANSHAHR	Holistic	2020
[20]	Statistical features, Support Vector Machine	IRANSHAHR	Holistic	2018
[28]	Connectionist Temporal Classification (CTC) , CNN	Sadri	Holistic	2020
[29]	Various CNN models	Sadri	Holistic	2020

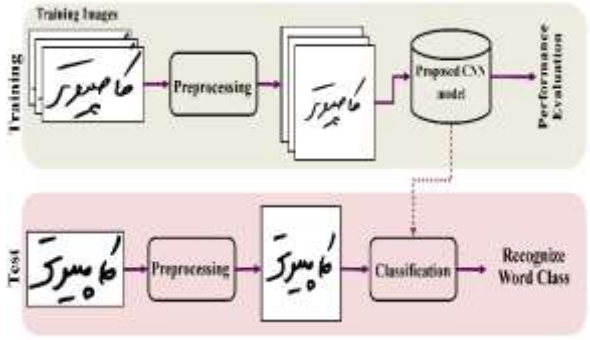


Figure 1. The general diagram for handwritten word recognition

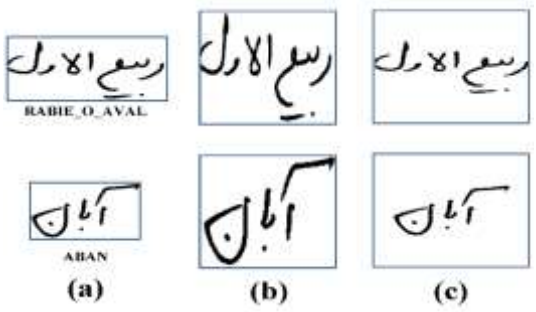


Figure 2. Left column(a): original image, Middle column (b): Resized image by normal scaling, and right column (c): Resized image by [32]

is utilized in most conventional approaches [1]. The end-to-end learning in the absence of handcrafted features is one of the remarkable characteristics of the CNN. Therefore in this paper, inspired by [1], a CNN model is adopted for holistic Persian off-line handwritten word recognition. As far as the authors are concerned, this is the first paper to focus on the effectiveness of CNN model for holistic recognition of Persian off-line handwritten. The architecture of the proposed CNN model is depicted in Figure 3. As can be seen, the proposed CNN networks have one input layer, five convolutional blocks with corresponding four max-pooling ($L_1 \cdots L_4$) and one average-pooling in the last block (L_5) for feature extraction, two fully-connected layers (L_6 and L_7) for classification, and finally one output layer. The details of the proposed architecture are listed in Table 2.

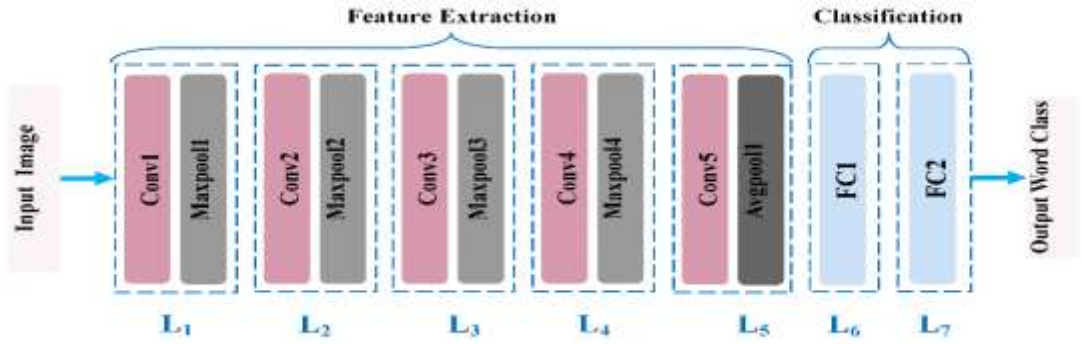


Figure 3. The proposed CNN model for Persian handwritten word recognition. This architecture contains seven layers ($L_1 \cdots L_7$) for extracting feature and classification

TABLE 2. The details of the proposed CNN model shown in Figure 3

Layer Name	Layer Type	No. of Filters	Kernel Size	Stride	Input Features	Output Features	No. of Parameters
L_1	Convolution (Conv1)	16	(7×7)	(1×1)	(1,224,224)	(16,218,218)	800
	Batch Normalization	---	---	---	(16,218,218)	(16,218,218)	32
	RELU	---	---	---	(16,218,218)	(16,218,218)	0
	Max Pooling (Maxpool1)	---	(2×2)	(2×2)	(16,218,218)	(16,109,109)	0
L_2	Convolution (Conv2)	32	(5×5)	(1×1)	(16,109,109)	(32,105,105)	12,832
	Batch Normalization	---	---	---	(32,105,105)	(32,105,105)	64
	RELU	---	---	---	(32,105,105)	(32,105,105)	0
	Max Pooling (Maxpool2)	---	(2×2)	(2×2)	(32,105,105)	(32,52,52)	0
L_3	Convolution (Conv3)	64	(3×3)	(1×1)	(32,52,52)	(64,50,50)	18,496
	Batch Normalization	---	---	---	(64,50,50)	(64,50,50)	128
	RELU	---	---	---	(64,50,50)	(64,50,50)	0
	Max Pooling (Maxpool3)	---	(2×2)	(2×2)	(64,50,50)	(64,25,25)	0
L_4	Convolution (Conv4)	64	(3×3)	(1×1)	(64,25,25)	(64,23,23)	36,928
	Batch Normalization	---	---	---	(64,23,23)	(64,23,23)	128
	RELU	---	---	---	(64,23,23)	(64,23,23)	0
	Max Pooling (Maxpool4)	---	(2×2)	(2×2)	(64,23,23)	(64,11,11)	0
L_5	Convolution (Conv5)	128	(3×3)	(1×1)	(64,11,11)	(128,9,9)	73,856
	Batch Normalization	---	---	---	(128,9,9)	(128,9,9)	256
	RELU	---	---	---	(128,9,9)	(128,9,9)	0
	Average Pooling (Avgpool1)	---	(2×2)	(2×2)	(128,9,9)	(128,2,2)	0
L_6	Fully Connected (FC1)	---	---	---	512	256	131,328
	RELU	---	---	---	256	256	0
L_7	Fully Connected (FC2)	---	---	---	256	125	32,125
Totally = 308,541							

According to Table 2, in general, the proposed CNN consists of 308,541 (≈0.3 million) parameters (weight) that must be trained in the training process. It is worth noting that the number of layers and characteristics of each layer have been adjusted experimentally. As shown in Figure 3, feature extraction and classification were conducted automatically. The main goal for the presented model design is to accomplish the best performance with the minimum number of layers.

4. EXPERIMENTAL RESULTS

This section explains experimental results after introducing the two datasets. All experiments were run on a machine with Intel® core i3 - 6300 CPU @3:70GHz, 16GB RAM, and NVidia® 1060Ti 6GB GPU. The experiments were implemented using PyTorch® framework installed on Microsoft® Windows 10. The back-propagation algorithm beside the Adam optimizeris were utilized to train the proposed CNN model. The proposed CNN model requires a number of hyper-parameters shown in Table 3. It should be noted that the hyper-parameter values were adjusted empirically.

4. 1. Dataset In the proposed method, two Persian handwritten word datasets including IRANSHAHR and Sadri was utilized. The Sadri dataset comprises text, dates and numbers, as well as words, numbers, signs, letters, and symbols. There are 62,500 words from 125 word classes in the Sadri dataset, which were collected randomly by 500 Persian authors, including 250 males and 250 females, of whom 10% were left-handed. IRANSHAHR is another dataset that contains 19,583 word images from 503 names of Iranian cities with approximately 38 sample images for each class. In this paper, to compare the proposed method with the state-of-the art, a subset of 200 out of the 503 city names was selected from IRANSHAHR. Figure 4 shows some samples of the two datasets.

4. 2. The First Experiment- Sadri Dataset In the first set of experiments, samples of Sadri dataset including 62,500 words were divided into three categories, 70% (0.7 × 62,500 = 43,750) for training, 15% (0.15 × 62,500 = 9, 735) for validation and 15% (0.15 × 62,500 = 9,735) for testing. Figure 5 shows validation accuracy and validation cost in different epochs. For comparative investigation, Figure 6 shows the confusion matrix of each category for Sadri dataset. The rows and columns denote the 10 first classes in Sadri dataset. The last column represents the rest of 125 classes in Sadri dataset.

TABLE 3. Hyper-parameter setting of the proposed CNN model

Hyper-parameter	value
Batch size	200
Number of epochs	100
Initial learning rate	0.001
L2regularization	0.001

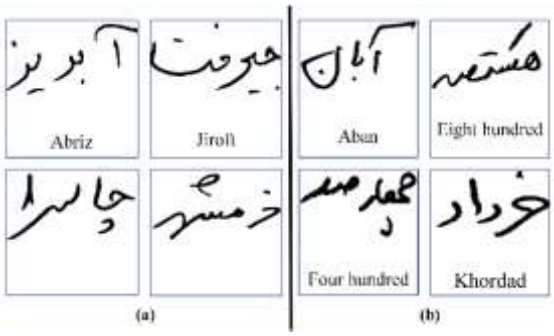


Figure 4. Several samples of the Persian handwritten word database, (a): IRANSHAHR dataset, (b) Sadri dataset

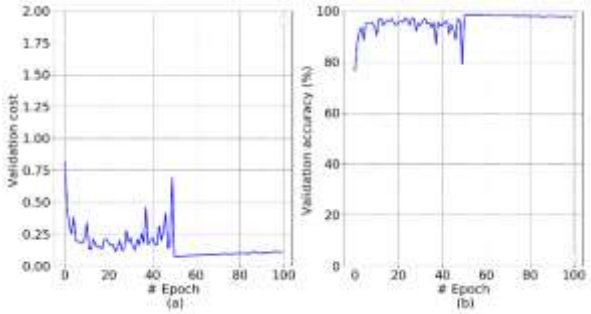


Figure 5. Performance of the proposed model in different epochs on Sadri dataset for validation set, (a) validation cost versus the number of epochs, (b) validation accuracy versus the number of epochs

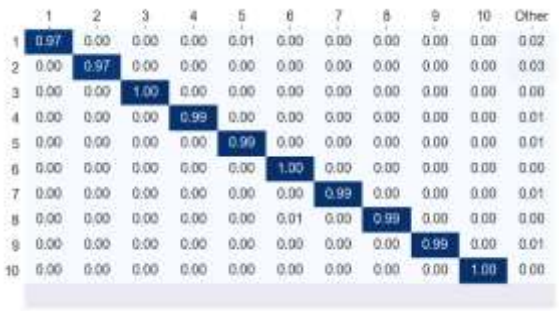


Figure 6. Confusion matrix where rows and columns represent the 10 first classes in each Sadri dataset. The last column represents the rest of 125 classes in Sadri dataset

4. 3. Effects of Rotation on Performance of the CNN Model This subsection attempts to assess stability of The presented CNN model in the presence of rotation errors. In this investigation, as shown in Figure 7, the test word samples are rotated in different angles (-15˚,-10˚,-5˚, 0˚, 5˚, 10˚, 15˚) and their corresponding recognition accuracies are depicted in Figure 8. As can be seen, rotation errors affect performance of the proposed model. In fact, this experiment shows the proposed model isn't rotation invariant.

4. 4. Effects of Pooling Types on the Proposed CNN Model The size of feature map in CNN is reduced by the pooling operation, which is invariant to image transformations [1]. Furthermore, research shows that pooling operations have a substantial effect on performance of the model [1]. The max pooling and average pooling are two pooling types widely used for designing CNN. Several studies [33] have sought to figure out the best one. In this paper, to design the proposed CNN, both pooling types have been utilized together: four max pooling operations in the first four convolution blocks as well as one average-pooling operation in the last convolution block (see Figure 3). Performance of the proposed model is evaluated once with only max pooling in all five convolution blocks, then with only average pooling in all five convolution blocks, and the results are listed in Table 4. As depicted in this table, compared to the average pooling and max-pooling strategies alone, the combined max-average pooling not only shows higher performance in terms of recognition accuracy, but also has fewer trainable parameters and training time.

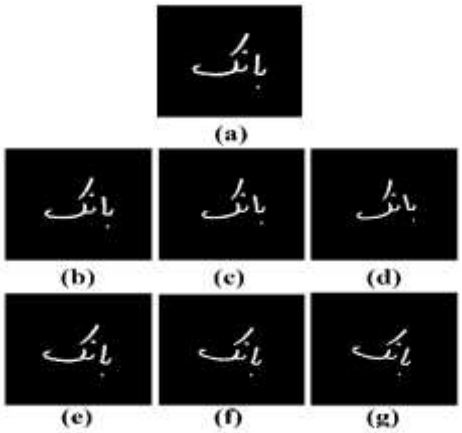


Figure 7. Different levels of rotation error (a) Original image, (b) Rotation angle = 5º, (c) Rotation angle = 10º, (d) Rotation angle = 15º, (e) Rotation angle = -5º, (f) Rotation angle = -10º, (g) Rotation angle = -15º respectively

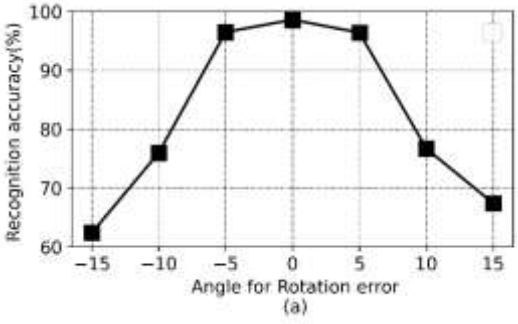


Figure 8. Performance evaluation for word images with various degrees of rotation

4. 5. The Second Experiment- Transfer Learning on IRANSHAHR Compared to the Sadri dataset, IRANSHAHR has smaller samples, approximately about 38 samples per class. In this subsection, two approaches are used. First, the proposed CNN model is trained from scratch on IRANSHAHR dataset, then the Transfer Learning (TL) approach is used [34]. TL aims to leverage knowledge from a related domain (called source domain) to improve learning performance in a target domain, when there is a small number of labeled training data in the target domain. In fact, due to the over-fitting phenomenon, it is not common to train the CNN model from scratch on a small dataset. Instead, the CNN model is trained on a richer dataset (i.e. Sadri), and then the trained model is fine-tuned on a small dataset (i.e. IRANSHAHR). So in this experiment, the trained CNN architecture on Sadri dataset was considered as the backbone network. Then, the backbone network was fine-tuned on IRANSHAHR by re-training only the last two layers (L_6 and L_7 in Figure 3). In the two experiments of this sub-section, samples of IRANSHAHR dataset are divided into three categories, 70% (0.7 × 19,583 = 13,708) for training, 15% (0.15×19,583 = 2,938) for validation and 15% (0.15×19,583 = 2,938) for testing. Figure 9 shows validation accuracy and validation cost in different epochs of the two approaches. As depicted in this figure, when training from scratch, the CNN network weights are randomly initialized, so the learning process shows a fluctuating behavior. In contrast, in the TL approach, CNN weights in the target domain are initialized based on the trained network's weights in the source domain. Thus, this initialization strategy triggers

TABLE 4. Evaluations of Different pooling types

Pooling Type	Training Time (minute)	# of Trainable Parameters	Test Accuracy (%)	
			Top 1	Top 5
Max-pooling	181.3	701,757	98.6	99.8
Average-pooling	183.4	308,541	97.9	99.8
Max-average pooling	**175.1**	**308.541**	**98.6**	**99.8**

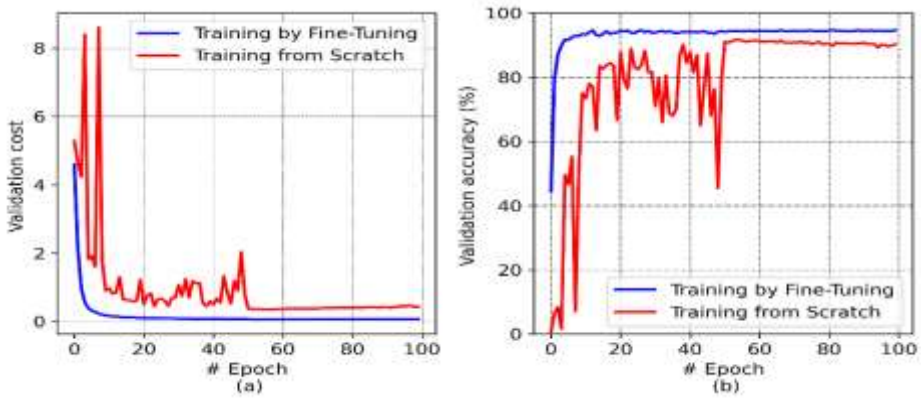


Figure 9. Performance of the proposed model in different epochs on IRANSHAHR dataset, (a): cost value versus the number of epochs, (b): validation accuracy versus the number of epochs for validation set

a smoother behavior in the learning process. This difference is clearly shown in Figure 9. As shown in Table 5, the TL approach outperforms other approach with regard to the recognition accuracy, test cost, and training time.

5. ANALYSIS OF ERROR

This section presents an analysis of errors on the proposed method. In this article, CNN model is proposed for the holistic Persian off-line handwritten word. The proposed CNN models have hieratical structures with many trainable kernels. The CNN model is essentially based on convolution operator, meaning that CNN kernels are convolved over the image for extracting meaningful information. While kernels in the primary layers extract common information like edge, line, etc., kernels in the later layers progressively extract more detailed and complex information from input sample images. For error analysis, first the average of images in each class is calculated by summing the images of that class divided by the total number of images in the class. Figure 10 shows the average image in class "آبان" (Aban). In fact, Figure 10 shows how people usually write the word "آبان" on paper.

On the other hand, the intensity frequency in average images of each class indicates that the kernels of a CNN will attach great importance to these regions during the training process. It means that a CNN learns the general structure of the images in each class during the training. However, during the test, the test images distant from the overall structure of its corresponding class will have a greater probability of error and vice versa. Based on this analysis, the errors in the test stage were split into two main groups on Sadri dataset. The first group contains errors occurring during data collection. Table 6 shows some of these errors. In the second group, as shown in Table 7, the test image samples are distant from average images in the corresponding class. It should be noted that

TABLE 5. Details of the two different approaches on IRANSHAHR dataset in the test phase

Database	Training Type	Test Accuracy (%)		Test Loss	Training Time (minute)
		Top1	Top5		
IRANSHAHR	From Scratch	90.8	98.1	0.39	22.3
	Fine-Tuning	94.6	99.0	0.24	10.4

Figure 10. Image average for the class "آبان"

this distance can be due to differences between the beginning or end of test image and the average structure of the corresponding class.

6. DISCUSSION AND COMPARISON

This section compares our study with the most recent state-of-the-art works on these two datasets in terms of

TABLE 6. The number of errors during the collection of Sadri dataset

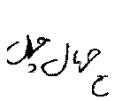
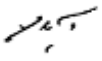
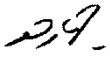
Word Image	Image Label in dataset	Word Image	Image Label in dataset
	پنج Five		چهل Forty
	چهل Forty		یازده Eleven

TABLE 7. The number of errors made by the proposed method on Sadri dataset

Word Image	Average Images in corresponding class	True Label	Predicted Label
		آبان ABAN	آذر AZAR
		چهارده Fourteen	چهارصد Four hundred
		تومان TOMAN	تلفن Phone
		خانم Miss	تمام End

recognition accuracy. To the best of our khowlwdge, few studies have explored CNN-based methods for the recognition of Persian handwritten off-line words.

Thus, motivated by [1], this paper adopted CNN architecture for holistic Persian handwritten word recognition. In addition, for the first time, this paper focuses on the effectiveness of CNN model for holistic Persian off-line handwritten word recognition. Compared to previous studies on IRANSHAHR dataset, the proposed model had a significantly higher improvement in recognition accuracy. Due to limited samples in each class of the IRANSHAHR dataset, all works had been selected from a limited subset of 503 classes for their analysis. In this paper, for the first time, we used the all 503 classes as shown in Table 8. As can be see, compared to IRANSHAHR, much fewer investigations have been conducted on Sadri dataset. As far as the authors are concerned, there are currently only studies based on Sadri dataset [28, 29]. The proposed method has several advantages over the existing methods. For example, the proposed method provides better results with fewer weights than the other methods. Moreover, compared to the wotk in literature [28], in which CNN was conducted for extracting a feature sequences and the RNN was used along with CTC for sequence labeling, the proposed method is end-to-end, meaning that feature extraction and classification are conducted automatically. Given the large size of trainable parameters, the method of Bonyani et al. [29] used data augmentation for data generation. In contrast, the proposed method eliminates the need for data augmentation due to efficient trainable parameters. In general the proposed holistic method has several critical advantages over other studies including:

- During the network training, appropriate features are extracted automatically. Hence, it eliminates the

TABLE 8. Comparing the performance of the proposed approach with the state-of-the-art methods

Dataset	Ref.	Method	Year	# of trainable parameters	# of classes	Accuracy (%)	
						Top1	Top5
IRANSHAHR	[15]	Self-Organization Map, discrete HMM	2001	-----	198	69.0	-----
	[11]	Image gradient, classifier fusion	2020	-----	200	89.0	96.4
	[16]	Fuzzy vector quantization, HMM	2001	-----	198	67.7	-----
	[10]	Right to Left, Left to Right HMM	2018	-----	200	84.4	97.9
	[17]	Lexicon reduction, discrete HMM	2008	-----	200	73.6	89.0
	[20]	Statistical features, Support Vector Machine	2018	-----	198	80.7	-----
	---	**Proposed Network**			**200**	**94.6**	**99.0**
				≈ 0.3M	**503**	**93.0**	**98.6**
Sadri	[28]	Connectionist Temporal Classification, bi-LSTM, CNN	**2020**	≈ 0.7M	125	98.8	99.3
	[29]	Various CNN models	**2020**	≈ 7M	125	98.8	-----
	---	**Proposed Network**	-----	**≈ 0.3M**	**125**	**98.6**	**99.9**

M : Million

manual feature extraction stage, which is a highly time-intensive and boring task in existing approaches.

- As the experimental results indicate, the presented method can be used as the backbone for other handwritten scripts with a reduced time complexity.

7. CONCLUSION AND FUTURE WORKS

In this article, an end-to-end method based on CNN architecture was adopted for holistic Persian off-line handwritten word recognition. To the best of our knowledge, this is the first paper to focus on the effectiveness of CNN model for holistic Persian off-line handwritten word recognition. For this purpose, two sets of experiments were carried out. In the first set of experiments, the presented method was assessed on the Sadri dataset. In the second experiment, two approaches were followed on IRANSHAHR dataset. In the first approach, the proposed CNN method was trained from scratch. In the second approach, the TL approach was adopted. The experimental results indicate that the presented method surpasses the state-of-the art in term of recognition accuracy. The error analysis was conducted on the Sadri dataset for the first time. In the future, the authors would like to extend their model to other scripts like Arabic, Latin, etc.

8. ACKNOWLEDGEMENT

We would like to appreciate Sadri et al. [14] for providing the dataset.

9. REFERENCES

1. Das, D., Nayak, D.R., Dash, R., Majhi, B. and Zhang, Y.-D., "H-wordnet: A holistic convolutional neural network approach for handwritten word recognition", ***IET Image Processing***, Vol. 14, No. 9, (2020), 1794-1805. DOI: 10.1049/iet-ipr.2019.1398.
2. Imani, Z., Ahmadyfard, Z. and Zohrevand, A., "Holistic farsi handwritten word recognition using gradient features", ***Journal of AI and Data Mining***, Vol. 4, No. 1, (2016), 19-25.DOI: 10.5829/idosi.JAIDM.2016.04.01.03.
3. Akbari, Y., Jalili, M.J., Sadri, J., Nouri, K., Siddiqi, I. and Djeddi, C., "A novel database for automatic processing of persian handwritten bank checks", ***Pattern Recognition***, Vol. 74, (2018), 253-265. DOI: https://doi.org/10.1016/j.patcog.2017.09.011.
4. Ye, M., Viola, P., Raghupathy, S., Sutanto, H. and Li, C., "Learning to group text lines and regions in freeform handwritten notes", in Ninth International Conference on Document Analysis and Recognition (ICDAR 2007). Vol. 1, (2007), 28-32. : 10.1109/ICDAR.2007.4378670.
5. Razzak, I., Kamran, I. and Naz, S., "Deep analysis of handwritten notes for early diagnosis of neurological disorders", in 2020 International Joint Conference on Neural Networks (IJCNN). (2020), 1-6. DOI: 10.1109/IJCNN48605.2020.9207087.
6. Vajda, S., Roy, K., Pal, U., Chaudhuri, B.B. and Belaid, A., "Automation of indian postal documents written in bangla and english", ***International Journal of Pattern Recognition and Artificial Intelligence***, Vol. 23, No. 08, (2009), 1599-1632. DOI: https://doi.org/10.1142/S0218001409007739.
7. Venu, G. and Xue, H., "Fast handwriting recognition for indexing historical documents", in First International Workshop on Document Image Analysis for Libraries, 2004. Proceedings. (2004), 314-320. DOI: 10.1109/DIAL.2004.1263260.
8. Sánchez, J.A., Romero, V., Toselli, A.H., Villegas, M. and Vidal, E., "A set of benchmarks for handwritten text recognition on historical documents", ***Pattern Recognition***, Vol. 94, (2019), 122-134. DOI: https://doi.org/10.1016/j.patcog.2019.05.025.
9. Bhowmik, S., Malakar, S., Sarkar, R., Basu, S., Kundu, M. and Nasipuri, M., "Off-line bangla handwritten word recognition: A holistic approach", ***Neural Computing and Applications***, Vol. 31, No. 10, (2019), 5783-5798. DOI: 10.1007/s00521-018-3389-1.
10. Abbaszadeh Arani, S.A.A., Kabir, E. and Ebrahimpour, R., "Combining rtl and ltr hmms to recognise handwritten farsi words of small-and medium-sized vocabularies", ***IET Computer Vision***, Vol. 12, No. 6, (2018), 925-932. DOI: 10.1049/iet-cvi.2017.0645.
11. Arani, S.A.A.A., Kabir, E. and Ebrahimpour, R., "Handwritten farsi word recognition using nn-based fusion of hmm classifiers with different types of features", ***International Journal of Image and Graphics***, Vol. 19, No. 01, (2019), 1950001. DOI: https://doi.org/10.1142/S0219467819500013.
12. Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S. and Lew, M.S., "Deep learning for visual understanding: A review", ***Neurocomputing***, Vol. 187, (2016), 27-48. DOI: https://doi.org/10.1016/j.neucom.2015.09.116.
13. Bayesteh, E., Ahmadifard, A. and Khosravi, H., "A lexicon reduction method based on clustering word images in offline farsi handwritten word recognition systems", in 2011 7th Iranian Conference on Machine Vision and Image Processing. (2011), 1-5. DOI: 10.1109/IranianMVIP.2011.6121550.
14. Sadri, J., Yeganehzad, M.R. and Saghi, J., "A novel comprehensive database for offline persian handwriting recognition", ***Pattern Recognition***, Vol. 60, (2016), 378-393.DOI: https://doi.org/10.1016/j.patcog.2016.03.024.
15. Dehghan, M., Faez, K., Ahmadi, M. and Shridhar, M., "Handwritten farsi (arabic) word recognition: A holistic approach using discrete hmm", ***Pattern Recognition***, Vol. 34, No. 5, (2001), 1057-1065. DOI: https://doi.org/10.1016/S0031-3203(00)00051-0.
16. Dehghan, M., Faez, K., Ahmadi, M. and Shridhar, M., "Unconstrained farsi handwritten word recognition using fuzzy vector quantization and hidden markov models", ***Pattern Recognition Letters***, Vol. 22, No. 2, (2001), 209-214. DOI: https://doi.org/10.1016/S0167-8655(00)00090-8.
17. Mozaffari, S., Faez, K., Märgner, V. and El-Abed, H., "Lexicon reduction using dots for off-line farsi/arabic handwritten word recognition", ***Pattern Recognition Letters***, Vol. 29, No. 6, (2008), 724-734. DOI: https://doi.org/10.1016/j.patrec.2007.11.009.
18. Broumandnia, A., Shanbehzadeh, J. and Rezakhah Varnoosfaderani, M., "Persian/arabic handwritten word recognition using m-band packet wavelet transform", ***Image and Vision Computing***, Vol. 26, No. 6, (2008), 829-842. DOI: https://doi.org/10.1016/j.imavis.2007.09.004.
19. Imani, Z., Ahmadyfard, A. and Zohrevand, A., "Introduction to database farsa: Digital image of handwritten farsi words (in persian)", in 11th Iranian Conference on Intelligent Systems in Persian, Tehran, Iran. (2013). DOI: https://civilica.com/doc/214715/.

20. Tavoli, R., Keyvanpour, M. and Mozaffari, S., "Statistical geometric components of straight lines (sgcsl) feature extraction method for offline arabic/persian handwritten words recognition", ***IET Image Processing***, Vol. 12, No. 9, (2018), 1606-1616.
21. Moghaddam, R.F., Cheriet, M., Adankon, M.M., Filonenko, K. and Wisnovsky, R., "Ibn sina: A database for research on processing and understanding of arabic manuscripts images", in Proceedings of the 9th IAPR International Workshop on Document Analysis Systems, Boston, Massachusetts, USA, Association for Computing Machinery. (2010 of Conference), 11–18.
22. Pechwitz, M., Maddouri, S.S., Märgner, V., Ellouze, N. and Amiri, H., "Ifn/enit-database of handwritten arabic words", in Proc. of CIFED, Citeseer. Vol. 2, (2002), 127-136.
23. LeCun, Y. and Bengio, Y., "Convolutional networks for images, speech, and time series", ***The Handbook of Brain Theory and Neural Networks***, Vol. 3361, No. 10, (1995), 1995.
24. Liu, L., Ouyang, W., Wang, X., Fieguth, P., Chen, J., Liu, X. and Pietikäinen, M., "Deep learning for generic object detection: A survey", ***International Journal of Computer Vision***, Vol. 128, No. 2, (2020), 261-318.DOI: 10.1007/s11263-019-01247-4.
25. Zohrevand, A., Imani, Z. and Ezoji, M., "Deep convolutional neural network for finger-knuckle-print recognition", ***International Journal of Engineering, Transactions A: Basics***, Vol. 34, No. 7, (2021), 1684-1693. DOI: 10.5829/ije.2021.34.07a.12.
26. Guo, G. and Zhang, N., "A survey on deep learning based face recognition", ***Computer Vision and Image Understanding***, Vol. 189, (2019), 102805. DOI: https://doi.org/10.1016/j.cviu.2019.102805.
27. Zohrevand, A., Sattari, M., Sadri, J., Imani, Z., Suen, C.Y. and Djeddi, C., "Comparison of persian handwritten digit recognition in three color modalities using deep neural networks, Cham, Springer International Publishing. (2020), 125-136. DOI: 10.1007/978-3-030-59830-3_11.
28. Safarzadeh, V.M. and Jafarzadeh, P., "Offline persian handwriting recognition with cnn and rnn-ctc", in 2020 25th International Computer Conference, Computer Society of Iran (CSICC). (2020), 1-10. DOI: 10.1109/CSICC49403.2020.9050073.
29. Bonyani, M., Jahangard, S. and Daneshmand, M., "Persian handwritten digit, character and word recognition using deep learning", ***International Journal on Document Analysis and Recognition (IJDAR)***, Vol. 24, No. 1, (2021), 133-143.DOI: 10.1007/s10032-021-00368-2.
30. Huang, G., Liu, Z., Van Der Maaten, L. and Weinberger, K.Q., "Densely connected convolutional networks", in Proceedings of the IEEE conference on computer vision and pattern recognition. (2017), 4700-4708.
31. Chollet, F., "Xception: Deep learning with depthwise separable convolutions", in Proceedings of the IEEE conference on computer vision and pattern recognition. (2017), 1251-1258.
32. Sabzi, R., Fotoohinya, Z., Khalili, A., Golzari, S., Salkhorde, Z., Behravesh, S. and Akbarpour, S., "Recognizing persian handwritten words using deep convolutional networks", in 2017 Artificial Intelligence and Signal Processing Conference (AISP). (2017), 85-90.DOI: 10.1109/AISP.2017.8324114.
33. Scherer, D., Müller, A. and Behnke, S., "Evaluation of pooling operations in convolutional architectures for object recognition", in Artificial Neural Networks – ICANN 2010, Berlin, Heidelberg, Springer Berlin Heidelberg. (2010), 92-101. DOI: https://doi.org/10.1007/978-3-642-15825-4_10.
34. Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H. and He, Q., "A comprehensive survey on transfer learning", ***Proceedings of the IEEE***, Vol. 109, No. 1, (2021), 43-76. DOI: 10.1109/JPROC.2020.3004555.

Persian Abstract

چکیده

به دلیل ماهیت پیوسته کلمات و تنوع زیاد نوشتاری در زبان فارسی، شکستن کلمات به زیر کلمات و حرف کار بسیار سختی است. روش‌های کلی‌نگر با در نظر گرفتن شکل کلی کلمه این قسمت را نادیده می گیرند. در این مقاله یک روش end-to-end و کلی‌نگر بر اساس شبکه های عصبی کانولوشنی عمیق برای بازشناسی کلمات دست‌نوشته فارسی به صورت برون خط ارائه شده است. مدل پیشنهادی تنها دارای پنج لایه کانولوشنی و دو لایه طبقه بندی است که منجر به کاهش محسوس در تعداد پارامترها می شود. در این مقاله همچنین تاثیر استراتژی های متفاوت برای pooling مورد مطالعه قرار گرفته است. هدف اصلی این مقاله معرفی یک روش جدید برای بازشناسی کلمات دست‌نویس فارسی است که در آن ویژگی‌ها به صورت خودکار استخراج می شوند. روش پیشنهادی روی دو پایگاه داده مشهور به نام ایرانشهر و صدری مورد ارزیابی قرار گرفت و به دقت بازشناسی معادل ۹۴.۶٪ روی ایرانشهر و ۹۸.۶٪ روی صدری رسید که نشان میدهد نسبت به روش های موجود کارائی بالاتری دارد.