VOL. 13, NO. 22, NOVEMBER 2018 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

KHMER HANDWRITTEN TEXT RECOGNITION WITH CONVOLUTION NEURAL NETWORKS

Bayram Annanurov[1,2] and Norliza Mohd Noor[1]
[1]Razak Faculty of Technology and Informatics, Universiti Teknologi Malaysia, Kuala Lumpur, Malaysia
[2]Computer Science Department, Zaman University, Phnom Penh, Cambodia
E-M ail: norliza@utm.my

ABSTRACT

This paper presents a novel pilot study on Khmer handwritten symbols recognition using Convolutional Neural Networks (CNNs). The motivation for this study is to develop a recognition system for digitizing large corpora of Khmer handwritten documents. Image data consists of six handwriting sample sets, each of which consists of 33 consonants (root radicals) and 17 vowels, giving a total of 561 syllables. A CNN-based model was trained for offline recognition of root radicals where one CNN was trained for recognition of a particular consonant giving 33 networks which were then combined into an assembly. The recognition results are compared against artificial neural network (ANN)-based classifier with full feature set and ANN -based classifier with dimensionality reduction. Feature correlation two-dimensional Fourier transformation (FT2D) and Gabor filters are used for dimensionality reduction. Recognition rate of Khmer handwriting (alpha syllabary system) is increased to 94.85% with Convolutional Neural Networks (CNN).

Keywords: handwritten text recognition, convolutional neural networks, classification, Khmer handwriting

INTRODUCTION

Pattern recognition aims to implement the ability of the human recognition in machines. Today computers can hardly match human abilities such as visual perception and understanding text. Handwriting recognition is the task of extracting text in graphical form into the symbolic representation. According to Surinta *et al.* [1], the aim of handwriting recognition is to let the recognizing system automatically identify the characters that are embedded in handwritten sample (manuscript).

A great variability of handwriting styles is the reason why handwritten text recognition is an area of interest. The challenges of the task are posed by numerous impacts like variety of writing systems, culture, occupation, medical condition, and many more [2-10].

Khmer (or Cambodian) is the language of the Khmer people and the official language of Cambodia. Being influenced by Sanskrit and Pali, it is one of the widely spoken languages in Southeast Asia. Symbols in Khmer are represented by one syllable and consist of a root for a consonant and a diactric for a vowel. There are 33 Khmer consonants that are commonly being used today. Some symbols still present in the alphabet but are not widely used. A set of dependent vowels is supplemented by some fourteen diactric symbols which also modify pronunciation of a consonant as well as the semantics of a syllable. Some consonants can be used as subscripts in special cases. Thus, there are more than 1000 different combinations that can be used in today's Khmer writing.

Several researchers have reported their work on Khmer and on other similar writing systems [6, 11-14]. Their objective has been to recognize printed text and, in some cases, handwritten text. A range of experiments is performed by Kheang *et al.*, Srun *et al.* towards analysis and recognition of printed Khmer characters and numerals [15-17]. The fact that most of the research up to date are based either on printed text and on a limited scope dataset (e.g. only numerals), motivates us to undertake this research.

Deep learning has also been used to recognize other scripts, including hieroglyphic as well as historical scripts. [10, 18-26].

Chao *et al.* [27] trained an unsupervised distributed Neural Net for feature learning that dramatically improved performance on large models. Also, distributing problems were addressed by Bagging-Down Stochastic Gradient Descent algorithm (SGD). The system featured the parameter server adding on the several model replicas, separated the updating and the training computing to accelerated the system. Some issues related to shared-memory learning were discussed in the paper.

The set of handwritten samples available for a research can be successfully expanded in order to increase the diversity and improve recognition [3]. Samples were automatically scaled, skewed and rotated in order to obtain more training data.

Analysis of writing with deep learning can be implemented in other numerous areas like diagnosis of Parkinson Disease [2], parallelizable deep learning model [4].

The current work differs from previously proposed methods both in the scope and modelling of the handwriting. It is designed to develop a recognition technique for a large collection of handwritten symbols with reduced space requirement. We evaluate the proposed method by comparing the recognition rate of the previous and current experiments. Our previous work includes Khmer handwritten text recognition with Artificial Neural Network and filtering [22] as well as comparison of various ways to reduce dimensionality [21].

A pilot study on Khmer handwriting recognition with CNN is presented in this paper as an attempt towards building a deep learning system. Performance is compared against other classifiers.

MATERIALS AND METHODS

Data collection

In this study, six sets of handwriting of the Khmer alphabets (33 consonants and 17 vowels) were obtained from six different donors who have deep knowledge of the Khmer language. The donors provided the alphabet samples by writing on a prepared table (33 x 17 blocks) on A3-size papers. The filled tables were scanned to obtain grayscale JPG images. Each sample was then isolated into an individual image forpre-processing.

Although there are hundreds possible distinct syllables, the scope of this study is limited to the most widely used consonants and vowels in Khmer alphabet.

The data set allows obtaining various samples of one consonant since each consonant is repeated within over 17 different syllables. Similarly, a vowel is repeated 33 times. This provides sufficient amount of data to construct a one-way classification system.

Samples of same consonant ក /k/ within different syllables are given in Figure-1. Samples of the same ក /ka/ syllable written by different people are shown in Figure-2.

Figure-1. Khmer Letter ក (/k/) with various diactrics written by the same person. [21, 22]

Figure-2. Khmer letter ក within the syllable (/ka/) written by different people. [21, 22]

Feature selection system

Pre-processing of the data set consists of normalization and feature selection (FS). Scanned samples were normalized and prepared for feature reduction. For benchmarking, we used artificial neural network (ANN)-based classifier with full feature set and an ANN-based classifier on data with reduced feature set. CNN is applied on the data with full feature set. The results are then evaluated based on the recognition rate.

Normalization

Samples were normalized in size and colour. For size normalization, white space was cropped off each sample, which was then resized to a standard size (32×32pixels). Java code is used due to ability of fast manipulation.

Colour normalization was based on a threshold, which was taken as 112. Grayscale pixels with colour value of less than 112 have been normalized to 0, and those over 112 is normalized to 225.

Correlation based feature selection

Feature set reduction helps speeding up the learning process and improve space requirement on the hardware. Reducing features can be done as far as it is possible to differentiate among classes. Otherwise, loss of information within the data set is imminent and the recognition will become impossible.

To do so, features that resemble each other are identified and only one is selected. Thus, the feature set F features is divided into subsets F_1, F_2, ..., F_N, so that all features in set F_i have high correlation and any two features in different subsets F_i and F_j have low correlation. Feature selection methods are briefly described in the next section. A more comprehensive description is given in our previous work [21, 22]. The results of feature selection are also used for benchmarking in current experiments.

Feature selection with fourier transformation

Two-dimensional Fourier Transform (FT2D) represents each sample in the frequency domain. For recognition purpose, a frequency is taken as a feature; and the feature set is reduced by selecting the frequencies with high amplitudes.

Gabor filters

Gabor filters are used to define an image in different frequency levels (frequency domain). Such representation yields a more thorough search of objects in a sample. Depending on the output of each Gabor filter, various features can be obtained and used in further analysis [23, 28].

Convolutional neural networks

Convolutional neural networks (CNN or ConvNet) is a mathematical tool that features invariance in shift, scale and distortion. LeCun *et al.* have described a LeNet-5, a model that successfully recognizes isolated characters [29]. Reported error rate for LeNet-5 were 0.8-0.96%. Even though the samples are different from Khmer alphabet, the model is a milestone in recognition technique.

CNN is also useful for analysis of other two-dimensional data [16, 30, 31]. Likewise, CNN shows high performance rates in tasks such as face recognition [19]. The structure of CNN is much more complex compared to classic setup such as Artificial Neural Nets.

A typical CNN uses small portions of a given image as input to the initial layer of its hierarchy. The information then progresses through the further hierarchical layers of the network where at each layer a filter is applied in order to obtain relevant features. The input image is convolved by small filters whose

coefficients are updated by training criteria. Hence, the first layer of CNN is essentially feature map resulting from of the convolution process, followed by a sub-sampling that reduces the dimensionality. Reducing dimensionality makes the system less prone to spatial shifts like rotation, skewing. This process can be repeated as many times as desired. The use of CNN for recognition of digits is described by LeCun *et al*. [29]. A typical topology is shown in Figure-3.

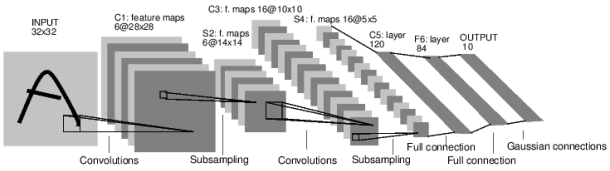

Figure-3. Architecture of a typical convolution neural network, here for digit recognition. [25, 29].

CNN can learn the underlying features that are not always visible to the naked eye. It does not explicitly extract features. On the other hand, the output of a layer can be taken as a representation of the input. Even the output layer, whose value predicts class, can be seen as a representation of current data sample. In fact, Razavian *et al.* [32] trained another classifier by using the information travelling between layers as features. The input size is also greatly decreased as well as computational cost. Current developments of hardware technology make it possible to implement greatly accelerated networks using publicly available libraries, such as TensorFlow [33], caffe[34], tfLearn[35] or Keras[36].

Building blocks of a Convolutional Neural Network have been extensively explained in many previous works [24-26, 28-30]. Typically, CNN consist of convolution layers, pooling layers, fully connected layers, learning rate, and activation function. Figure-3 summarizes the whole process.

Experimental setup

As mentioned in the data collection section, each of the six datasets is comprised of 561 (33X17) symbols. Division of each image is filed into an individual image of each syllable, cropping white space, and resizing to standard size are implemented programmatically in Java. Pre-processing, feature selection and basic classification were completed in previous experimentations [23,24]. Figure 4 shows the Block diagram of handwritten Khmer characters recognition process with FT2D and Neural Net.

For CNN setup we use two convolution layers, each followed by a pooling step. ReLu is used as activation function in fully connected layer, since it showed more significant results than other conventional functions like sigmoid or arctan.

Initially, multiclass classification, where the network would recognize one class out of 33 has been implemented. The recognition rate in that case turned out to be low, failing to yield any decisive outcome.

We built the CNN to recognize one consonant out of all (yes-no classification). For each dataset, we train 33 different CNN networks. At each training, one consonant is assumed as positive class, and all other – as negative. For example, at each run, the system is recognizing if the input is a sample of the letter "ក (/k/)". In particular /ka/ is of class "yes", unlike other consonants. The overall model is taken as the assembly of all 33 networks.

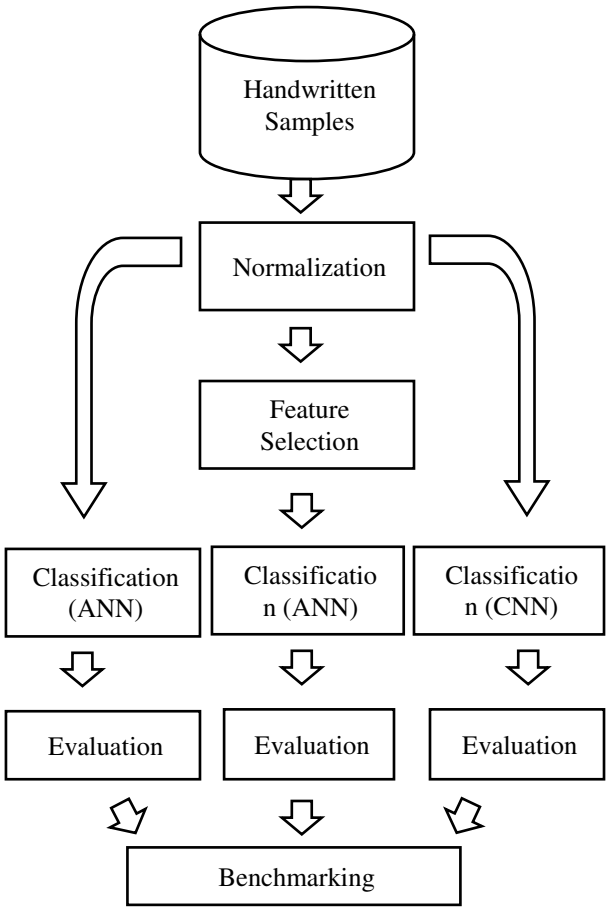

Figure-4. Block diagram of handwritten Khmer characters recognition process.

RESULTS AND DISCUSSIONS

Current experiments show that CNN yields best recognition rates. In order to assess the predictive performance of the model, we combine all data sets and perform CNN-based classification with k-fold cross-validation, where k = 2,3,5,10.

Recognition rates are averaged in order to obtain the recognition for CNN over that particular database. Results of recognition for Khmer handwriting are shown in Table-1.

During k-fold cross-validation, data is divided into k parts (folds). All parts but one (k-1) are used for training the model. When the model is trained it is applied to the remaining fold, and the recognition rate is recorded. Folds are then rotated, model is reset. Entire procedure is repeated$k$ times. Every time a different fold is kept for validation, while the model is retrained on remaining

folds. As a result, the recognition rate of the model is taken as the average of the obtained *k* recognition rates.

Table-1. Recognition rates under various settings.

	Recognition rate (%)				
Data set	**Full Feature Set**	**Corre -lation based**	**FT2D**	**Gabor filters**	**CNN based**
1	7.7	7.5	26.49	24.82	93.5
2	14.0	15.1	16.81	15.78	94.7
3	15.0	20.1	22.67	30.13	97.3
4	14.0	17.1	6.45	29.41	93.8
5	9.8	6.9	19.75	22.81	95.9
6	9.3	7.1	17.79	30.70	93.9
Ave-rage	11.63	12.3	18.33	25.61	94.85

Table-2 summarizes the recognition rates. The classifier has been trained a total of 20 (2+3+5+10) times. Each time 33 different convolutional networks have been trained and their performance has been averaged.Distance based classification based on full feature set, as well as classification based on FS with Gabor filters, FT2D and feature correlation has not shown very high improvement oven classification on full feature set. Convolutional Neural Net (CNN), however, greatly improved classification. With CNN, recognition rate increased up to 93.5-97.3%. CNN also performs very well on the combined data set, where samples collected from all six donors are consolidated. CNN has been trained with k-Fold validation for various values of k (k=2,3,5,10) to ensure generalization and avoid overfitting.

Table-3 shows the comparison of our proposed method with previous work on Khmer Handwriting Recognition. Srun *et al.* achieved higher recognition accuracy for Khmer printed and scanned text using ANN [17]. Our proposed method can achieve similar high recognition accuracy of 94.85% for Khmer handwritten text using CNN.

Table-2. k-fold validation for CNN-based recognition on combined data

k (folds)	**k = 2**	**k = 3**	**k = 5**	**k = 10**
Recognition Rate (%)	99.1	92.7	93.3	92.0

CONCLUSIONS AND FUTURE WORK

This paper proposes CNN model for offline Khmer handwriting recognition. The CNN-based recognition system introduced here greatly outperforms the earlier classifiers, even with the reduced feature sets. In future, it is possible to reduce the features by selecting or combining batches of pixels, or by any other transformation. The main criteria for a text recognition system is speed and precision. The speed of recognition is not taken in consideration in this work and is left for future work.

The experiments described here are limited to Khmer consonants. Work on vowels, numerals, and other symbols is yet to be performed. The current work on Khmer handwriting recognition is very limited. Research in a broader scope has yet to be conducted. This attempt is the first to use CNN and the first to use a wide range of handwritten symbols for Khmer handwriting recognition.

Table-3. Comparison of our proposed method with previous work on khmer handwriting recognition.

Work	**Dataset**	**Data**	**Number of samples**	**Classifier**	**Average accuracy**
Ye *et al.*[6]	Digital data (PC mouse, stylus pen)	Khmer, Myanmar Characters	242	Stock methods (GraffitiTM, EdgeWrite)	Writing pace
Sok and Taing[13]	Printed and scanned text	Khmer Characters	3000	SVM	98%
Meng and Morariu [14]	Printed and scanned text	Khmer Characters	215	ANN	65%
Thumwarin *et al.*[11]	Scanned text	Khmer letters and digits	6750	Distance-Based	98%
Kruy and Kameyama[15]	Printed and scanned text	Khmer words	1104	SIFT + distance-based	98%
Kheang *et al.* [16]	Printed and scanned text	Khmer words	110713	WFST	~73%
Srun *et al.* [37]	Printed and scanned text	Khmer Characters	33	ANN	97%
Proposed method	Handwritten Text	Khmer Characters	3366	CNN	94.85%

REFERENCES

[1] Surinta O, Karaaba MF, Schomaker LRB. and Wiering MA. 2015. Recognition of handwritten characters using local gradient feature descriptors. Engineering Applications of Artificial Intelligence, 45:405-414. DOI: 10.1016/j.engappai.2015.07.017.

[2] Pereira CR, Weber SAT. and Hook V. 2016. Deep learning-aided Parkinson's disease diagnosis from handwritten dynamics. Proc. SIBGRAPI, Sao Paulo, Brazil.

[3] Song X, Gao X, Ding Y. and Wang Z. 2016. A handwritten Chinese characters recognition method based on sample set expansion and CNN. Proc. ICSAI Shanghai, China.

[4] Angelov P. and Gu X. 2017. MICE: Multi-layer multi-model images classifier ensemble. Proc CYBCONF, Exeter, UK.

[5] Calderón A, Roa S. and Victorino J. 2003. Handwritten digit recognition using GCNNs, Proc. CIIC, Bogotá, Colombia.

[6] Ye KT, Phavy O. and Urano Y. 2008. Positional gesture for advanced smart terminals: Simple gesture text input for syllabic scripts like Myanmar, Khmer and Bangla. Proc. K-INGN, Geneva, Switzerland.

[7] Lee H. and Verma B. 2012. Binary segmentation algorithm for English cursive handwriting recognition. Pattern Recognition, 45(4):1306-1317. DOI: 10.1016/j.patcog.2011.09. 015.

[8] Zamora-Martínez F, Frinken V, España-Boquera S, Castro-Bleda MJ, Fischer A. and Bunke H. 2014. Neural network language models for off-line handwriting recognition. Pattern Recognition, 47(4): 1642-1652. DOI 10.1016/j.patcog.2013.10.020.

[9] Kira K. and Rendell LA. 1992. Feature selection problem: traditional methods and a new algorithm. Machine Learning Proceedings, pp 249–256, AAAI. DOI 10.1016/B978-1-55860-247-2.50037-1.

[10] Wicht B, Fischer A. and Hennebert J. 2016. Keyword spotting with convolutional deep belief networks and dynamic time warping. Proc. ICANN, Barcelona, Spain. pp. 113-120.

[11] Thumwarin P, Khem S, Janchitraponvej K. and Matsuura T. 2008. On-line writer dependent character recognition for Khmer based on FIR system characterizing handwriting motion. Proc. ECTI-CON, Krabi, Thailand.

[12] Wang WL. and Tang MH. 2015. A normalization process to standardize handwriting data collected from multiple resources for recognition. Procedia Computer Science, 61:402-409. DOI: 10.1016/j.procs.2015.09.171.

[13] Sok P. and Taing N. 2014. Support vector machine (SVM) based classifier for Khmer printed character-set recognition. Proc. APSIPA. Siem Reap, Cambodia.

[14] Meng H. and Morariu D. 2014. Khmer character recognition using artificial neural network. Proc. APSIPA, Siem Reap, Cambodia.

[15] Kruy V. and Kameyama W. 2016. Preliminary experiment on Khmer OCR. Proc. FIT, Tokyo, Japan.

[16] Kheang S, Katsurada K, Iribe Y. and Nitta,T. 2016. Using reversed sequences and grapheme generation rules to extend the feasibility of a phoneme transition network-based grapheme-to-phoneme conversion. Proc. IEICE. pp. 1182-1192.

[17] SrunS. 2015. Applying Tesseract for Khmer optical character recognition. ASEAN-UEC Symposium.

[18] Sas J. and U. Markowska-Kaczmar U. 2012. Similarity-based training set acquisition for continuous handwriting recognition. Information Sciences, 191:226-244.DOI: 10.1016/j.ins.2011.12.023.

[19] Sermanet P, Kavukcuoglu K. and LeCun Y. 2009. EBLearn: open-source energy-based learning in C++. Proc. ICTAI, Newark, NJ, USA.

[20] Richarz J, Vajda S, Grzeszick R. and Fink GA. 2014. Semi-supervised learning for character recognition in historical archive documents. Pattern Recognition, 47(3):1011-1020.DOI: 10.1016/j.patcog.2013.07.013.

[21] Annanurov B. and Noor NM. 2017. Feature selection for Khmer handwritten text recognition. Proc. ElConRus, St. Petersburg, Russia.

[22] Annanurov B. and Noor NM. 2016. Handwritten Khmer text recognition. Proc. WIECON-ECE, Pune, India, pp. 176-179.

[23] Hafemann LG, Sabourin R. and Oliveira LS. 2016. Writer-independent feature learning for offline

signature verification using deep convolutional neural networks.Proc. IJCNN, Vancouver, BC, Canada.

[24] Zhang XY, Yin F, Zhang YM, Liu CL. and Bengio Y. 2018. Drawing and Recognizing Chinese Characters with recurrent neural network. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(4): 849-862. DOI: 10.1109/TPAMI.2017.2695539.

[25] Liu Q, Tan T. and Yu K. 2016. An investigation on deep learning with beta stabilizer. Proc. ICSP, Búzios, Brazil.

[26] Saabni R. 2016. Recognizing handwritten single digits and digit strings using deep architecture of neural networks. Proc. AIPR, Lodz, Poland.

[27] Chao Q, Gao XG. and Chen D. 2016. On distributed deep network for processing large-scale sets of complex data. Proc. IHMSC, Hangzhou, China.

[28] Kamarainen JK. 2012. Gabor features in image analysis. Proc. IPTA, Istanbul, Turkey.

[29] LeCun Y, Bottou L. and Haffner P. 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278-2323. DOI: 10.1109/5.726791.

[30] Simonyan K. and Zisserman A. 2014. Two-stream convolutional networks for action recognition in videos. Proc., NIPS, Montréal, Canada.

[31] He K, Zhang X, Ren S. and Sun J. 2016. Deep residual learning for image recognition. Proc. CVPR, Las Vegas, NV, USA.

[32] Razavian AS, Azizpour H, Sullivan J. and Carlsson S. 2014. CNN features off-the-shelf: An astounding baseline for recognition. Proc. CVPRW, Columbus, OH, USA.

[33] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, Kudlur M, Levenberg J, Monga R, Moore S, Murray DG, Steiner B, Tucker P, Vasudevan V, Warden P, Wicke M, Yu Y. and Zheng X. 2016. TensorFlow: a system for large-scale machine learning. Proc. OSDI, Savannah, GA, United States. pp. 265-283.

[34] Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S. and Darrell T. 2014. Caffe: convolutional architecture for fast feature embedding. Proc. MM'14, Orlando, Florida, USA. pp. 675-678.

[35] Tang Y. 2016. TF.Learn: TensorFlow's high-level module for distributed machine learning. arXiv preprint arXiv:1612.04251.

[36] Charles PWD.2013. Keras. GitHub repository, https://github.com/keras-team/keras.

[37] Srun S. 2011. Applying backpropagation for Khmer printing character recognition. Proceedings of Japan-Cambodia Joint Symposium on Information Systems and Communication Technology, Pnom-Penh, Cambodia. pp. 135-136.