Indonesian Journal of Electrical Engineering and Computer Science
Vol. 15, No. 3, September 2019, pp. 1322~1328
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v15.i3.pp1322-1328

Comparison of convolutional neural network and bag of features for multi-font digit recognition

Nasibah Husna Mohd Kadir, Sharifah Nur Syafiqah Mohd Nur Hidayah, NorasiahMohammad, Zaidah Ibrahim
Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Malaysia

Article Info

Article history:

Received Nov 1, 2018
Revised Mar 10, 2019
Accepted Apr 25, 2019

Keywords:

Bag of features (BoF)
CNN
Deep learning
Font digit recognition

ABSTRACT

This paper evaluates the recognition performance of Convolutional Neural Network (CNN) and Bag of Features (BoF) for multiple font digit recognition. Font digit recognition is part of character recognition that is used to translate images from many document-input tasks such as handwritten, typewritten and printed text. BoF is a popular machine learning method while CNN is a popular deep learning method. Experiments were performed by applying BoF with Speeded-up Robust Feature (SURF) and Support Vector Machine (SVM) classifier and compared with CNN on Chars74K dataset. The recognition accuracy produced by BoF is just slightly lower than CNN where the accuracy of CNN is 0.96 while the accuracy of BoF is 0.94.

Corresponding Author:

Zaidah Ibrahim,
Faculty of Computer and Mathematical Sciences,
Universiti Teknologi MARA,
Shah Alam, Selangor, Malaysia.
Email: zaidah@tmsk.uitm.edu.my

1. INTRODUCTION

Font digit recognition is an improvement of naturally perceiving handwritten character and numerals by machine or personal computer and it is also a part of character recognition technology. This improvement can be connected to program recognizable proof of computer character information for example in medical documentation [1], bank check processing and postal mail sorting [2]. Since there exists various digitized characters, it is difficult to effectively distinguish a great deal of computer written character figures because of the thousands of numerals font types. With the quick advancement of worldwide data and the expanding request of technology, the use of digitized character advanced acknowledgment is earnest. Various techniques have been explored for font digit recognition such as Nearest Neighbor Calculation [2] and Neural Network [3]. The scientific capacity of complex grouping issue and the speculation capacity of systems are restricted, and high acknowledgment exactness cannot be accomplished. But with the advancement of deep learning technology and the rise of the Convolution Neural Network (CNNs) gives the likelihood to take care of this issue.

CNN is part of deep learning that utilizes mostly to group images, cluster them by likeness, and perform picture acknowledgment within the scenes. CNN has shown amazing enhancement in various object recognition such as iris recognition [4], traffic sign recognition [5-6], face recognition [7-9], fruit recognition [10], leaf recognition [11] and font recognition [12]. The common structure of a CNN comprises of layers of neurons. A neuron takes input values, does computations and passes the result to the next layer. Each model has its own layers of convolution and computational complexity. Modifications can be made to the architecture of CNN to increase its performance by applying bi-linear CNN [13] and multi-maxpooling layers [14].

Apart from CNN, another popular technique for object recognition is Bag of Features (BoF). BoF is a machine learning technique that represents an image as orderless collections of nearby highlights [15]. The term BoF comes from the term Bag of Words (BoW) utilizes in textual information recovery. BoF with Speeded-Up Robust Features (SURF) with Support Vector Machine (SVM) classifier have been applied for food image recognition [15], face recognition [16], text classification [17], and leaf recognition [18] with good results.

A comparative study between CNN and BoF has been performed for leaf recognition which indicates that BoF is better than CNN [19] whereas CNN is better than BoF for fruit recognition [20]. Due to the inconsistencies in the accuracy performance of CNN and BoF, this paper conducts a comparative study between CNN and BoF for multiple font digit recognition. This paper is organized as follows. The next section explains about the overall architecture of CNN and BoF, followed by discussions on results and analysis of the accuracy performance. The last section concludes this paper and states the future work.

2. OVERALL ARCHITECTURE

2.1. Convolutional Neural Network (CNN)

The dataset follows the imperatives given by the CNN models. One of the imperatives is with respect to the measure of the individual image. In this work, the CNN and BoF utilize an exact image size for training and testing. Other than the imperatives from the models, the equipment utilized plays an imperative portion on the execution. So, to test this CNN and BoF models, a high-performance laptop is utilized to conduct the test. The laptop used has 8 gigabytes of Random-Access Memory (RAM).

The dataset used for this experiment is Chars74K dataset [21]. It comprises of images of handwritten characters and digits. Each of the character comprises of 55 varieties of indistinguishable images. Since in general data preparation takes a long time, this work covers a transcribed font digits as it were, which are numbers 0 to 9. This makes the full number of images utilized within the test is 1100. The images were resized to 227 x 227 pixels.

Figure 1 shows an illustration of CNN architecture developed in this research. It starts with resized the images to 277x227, utilization composed of convolve, maximum pooling and ReLu layers, classification layer and lastly the digit is classified. This project experiments with different combination of convolve layer, pooling and ReLu layers to examine the effect of the number of layers to the digit recognition accuracy. The multiple font digit recognition models has maximum of 3 layers including five convolutional layers, three maxpooling or down-sampling layers and one classification layer. In the CNN, the step consistently convolutes the input data with multiple, different filters to extract lineaments then the subsampling layer summarizes the detected features into a features map [22].

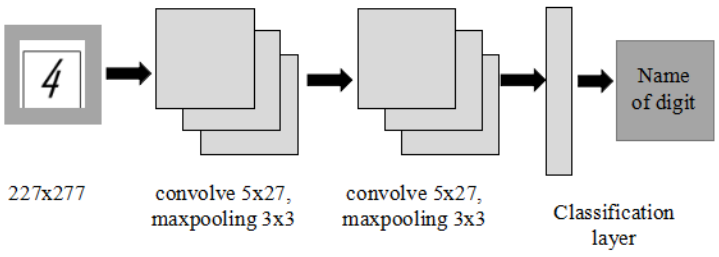

Figure 1. An illustration of CNN architecture

2.2. Bag of Features (BoF)

The Bag of Features (BOF) is one of the machine learning techniques often used in computer vision. It is also known as Bag of Visual Words [23]. This method is constructed from unstructured collections of independent visual features which come from the process of image extraction. An image can be transformed as a vector of features. The image is represented as a histogram of codes.

Figure 2 shows the general framework of BoF that involves two phases [23]. The first phase is handled with SURF features. SURF contains interest point detector which locates the significant points in the image and descriptor which describes the features of the significant points and features construction of the interest points [24]. Second phase is encoding and pooling. Encoding is about transforming the features into

local according to predefined codebook based on algorithm of the training samples while pooling is implementing features into global representation.

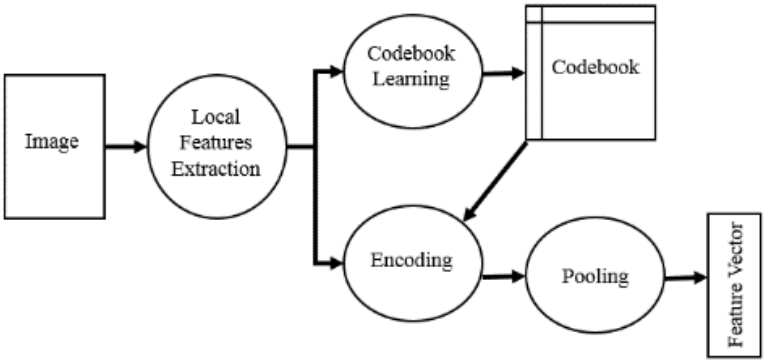

Figure 2. BoF frameworks [23]

3. RESULTS AND ANALYSIS

MATLAB is used for the experiment of multi-font digit recognition and Chars74k dataset [21] is used for training and testing.

3.1. Dataset

The Chars74k has been widely used and known as character recognition benchmark [21]. In the dataset, symbols used in both English and Kannada are available. In the English language, Latin script (excluding accents) and Hindu-Arabic numerals are used. The images are normalized and centered by center of mass in 28x28 fields. The images contain grey levels as a result of the anti-aliasing technique. This dataset contains 10 classes of digits which are 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9. It consists of 770 training images and 330 test images where each class has exactly 55 images. All of the images for each class are represented in different views to add robustness to the technique applied. The size of the images for each class is 128x128 pixels and all the images were resized to 227x227 for this experiment to fit into the Matlab program for CNN and BoF. Figure 3 shows some sample images from Chars74K dataset.

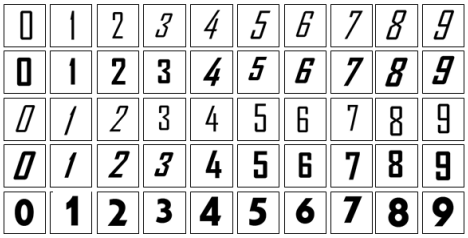

Figure 3. Sample images from Chars74K dataset [21]

3.2. Convolutional Neural Network (CNN)

As depicted in Figure 1, our CNN is composed of 2 layers of convolve, ReLu and pooling. The CNN takes a gray scale image (one channel) as input similar as the work in [24]. The size of the image in the input layer is 227x227x1 pixel. Each convolutional layer convolves the output of its previous layer with a set of learned kernels, followed by the ReLu, and max-pooling layer. This makes convolutional networks computationally capable, grant them to extent to large images when the exquisite transformation can be implemented as a distinct convolution rather than a fully general matrix multiplication [25].

Table 1 lists the accuracy performance of CNN for multi-font digit recognition. Different combination of values of the kernel size for both convolve layers are performed to determine the best accuracy results. As depicted in Table 1, Chars74k datasets is tested ten times to obtain the best result based

on different convolve layer and learning rate. For gray scale images, the first layer and second layer use the same value for the convolve layer (5,27) and (5, 27), the accuracy is 0.9633% with total training and testing time of 5 minutes and 19 seconds. Figure 4 shows the Visual Graph for CNN of training and validation of data for this experiment. This is the best results for the digit recognition that produce the best performance that is 0.9633%.

On the other hand, the accuracy result decreases drastically when the CNN is experimented with one and three layers of convolve, ReLu and pooling. Results illustrated in Table 1 indicate that different number of layers for CNN arrive to different results. Furthermore, the higher is the number of layers does not really produce better accuracy results. In this case, one layer of convolve, ReLu and pooling only extracts the low-level features that are the edges which is insufficient for multi-font digit recognition. In the case of three layers of convolve, ReLu and pooling, it obtains low accuracy since the data has been compressed so much after three downsized processes by the max-poling layer that the data does not really represent the digits anymore.

Table 1. Accuracy Performance of CNN for Multi-Font Digit Recognition

Input	*No of layers*	*Convolve Layer*	*Stride in maxpool*	*Epoch, Learning Rate*	*Accuracy*
Grayscale Image	Single set of layer	5, 20	3	10, 0.0001	0.3167
		4, 20	3	10, 0.0001	0.2300
		9, 16	3	10, 0.001	0.1000
		9, 16	3	10, 0.0001	0.1200
		9, 40	3	10, 0.0001	0.1467
	Two sets of layers	5, 15 5, 15	3 3	10, 0.0001	0.8933
		5, 25 5, 25	3 3	10, 0.0001	0.9033
		5, 27 5, 27	3 3	10, 0.0001	0.9633
		5, 26 5, 26	3 3	10, 0.0001	0.9200
		5, 27 5.27	3 3	10, 0.001	0.7233
	Three sets of layers	5, 20 5,20 5,20	3 3 3	10, 0.0001	0.2500

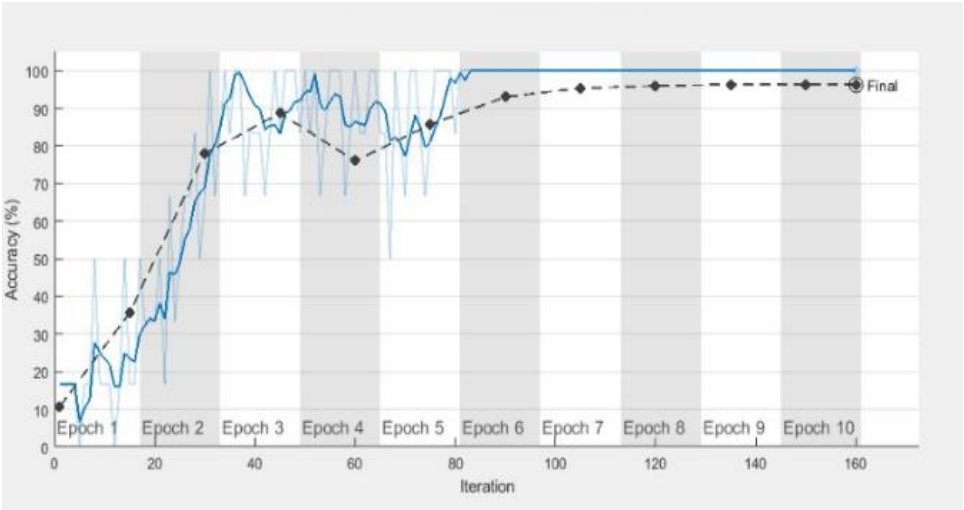

Figure 4. Visual Graph for CNN with two sets of layers of convolve, ReLu and maxpooling for multi-font digit recognition

3.3. Bag of Features (BoF)

The deployed BoF environment shown in this experiment includes the use of the Speeded up Robust Features (SURF) technique as a feature extractor and image encoder. For clustering, K-means algorithm is used. The strongest features are kept to 80 percent from each category. A vector quantization technique maps key-points from every training image is mapped into a unified dimensional histogram vector (Bag-of-Features) after K-means clustering. This histogram acts as an input vector for SVM classifier to build the training set. In the testing phase, the key-points are extracted and fed into the cluster model to map them into a BoF vector, which is finally fed into SVM training classifier to recognize the testing image. Based on the experiment, the average accuracy is 0.94. Figure 5 shows the visual word graph for BoF of the testing data for this experiment.

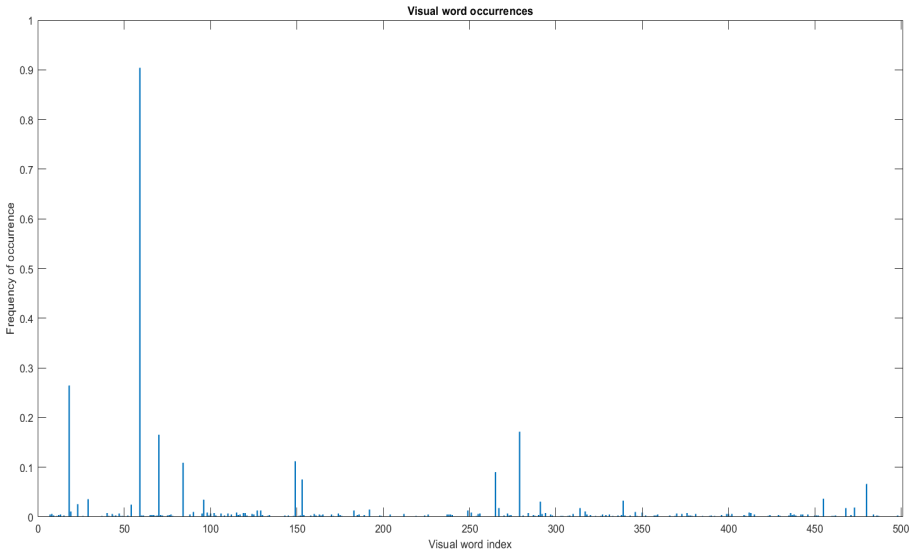

Figure 5. Visual Word Graph for BoF

4. CONCLUSION

In this research, different layers of CNN has been investigated to determine the optimum number of layers for multi-font digit recognition. The higher the number of layers does not really guarantee the achievement of a high accuracy. The performance of the CNN depends on the data itself, besides the number of convolve, ReLu and pooling layers, it also depends on the size of the filter image, epoch and learning rate. Smaller learning rate slows down the training process but may reach to better accuracy. BoF produces slightly lower accuracy compared to CNN. This shows that BoF can still be considered as a strong method for multi-font digit recognition based on the accuracy performance. Experimental results show that the CNN perform slightly better than BoF. In future work, more evaluations will be performed on CNN parameters for other datasets and different types of CNN architecture. Besides that, other features and classifiers will be investigated for BoF to compare the accuracy performance between machine learning and deep learning.

ACKNOWLEDGEMENTS

The authors would like to thank the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, for sponsoring this research.

REFERENCES

[1] Jia, L., & Sun, Y. (2018, May). "Digital Recognition based on Improved LENET Convolution Neural Network". in *Proceedings of the 2018 International Conference on Machine Learning Technologies*, pp. 24-28.

[2] Babu, U. R., Chintha, A. K., &Venkateswarlu, Y. (2014). "Handwritten Digit Recognition using Structural, Statistical Features and K-Nearest Neighbor Classifier". *International Journal Information Engineering and Electronic Business*, vol. 6(1), pp. 62-68.

[3] Md Saufi, M., Zamanhuri, M. A., Mohammad, N. and Ibrahim, Z. (2018). "Deep Learning for Roman Handwritten Character Recognition", *International Journal of Electrical Engineering and Computer Science (IJEECS)*, Vol. 12, No. 2, pp. 455-460.
[4] Nguyen, K., Fookies, C., Ross, A. and Sridharan, S. (2017). "Iris Recognition With Off-the-Shelf CNN Features: A Deep Learning Perspective". *IEEE Access* (Vol. 6) Visual Surveillance Biometrics: Practices, Challenges, and Possibilities.
[5] Shustanov, A. and Yakimov, P. (2017). "CNN Design for Real-Time Traffic Sign Recognition". *Procedia Engineering* 201, pp. 718-725.
[6] Luo, H., Yang, Y., Tong, B., Wu, F. and Fan, B. (2018). "Traffic Sign Recognition using A Multi-Task Convolutional Neural Network". *IEEE Transactions on Intelligent Transportation Systems*, Volume 19, Issue 4, pp. 1100-1111.
[7] Yin, X. and Liu, X. (2018). "Multi-task Convolutional Neural Network for Pose-Invariant Face Recognition". *IEEE Transactions on Image Processing*, Volume 27, Issue 2, pp. 964-975.
[8] Farfade, S. S., Saberian, M. J. and Li, L. J. (2015). "Multi-view Face Detection using Deep Convolutional Neural Networks". *Proceedings of the 5th ACM on International Conference on Multimedia Retrieval*, pp. 643-650.
[9] Nur Ateqah binti Mat Kasim, Nur Hidayah Binti Abd Rahman, Zaidah Ibrahim and Nur Nabilah Abu Mangshor (2018). "Celebrity Face Recognition using Deep Learning". *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, Vol. 12, No. 2, November 2018, pp. 476-481.
[10] Nur Azida Muhammad, Amelina Ab Nasir, Zaidah Ibrahim and Nurbaity Sabri (2018). "Evaluation of CNN, AlexNet and GoogleNet for Fruit Recognition". *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, Vol. 12, No. 2, November 2018, pp. 468-475.
[11] Nurbaity Sabri, Zalilah Abd Aziz, Zaidah Ibrahim, Muhammad Akmal Rasydan bin Mohd Rosni and Abdul Hafiz bin Abd Ghapul (2018). "Comparing Convolutional Neural Network Models for Leaf Recognition". *International Journal of Engineering Technology*, 7 (3.15) (2018) pp. 141-144
[12] Tensmeyer, C., Saunders, D., & Martinez, T. (2017, November). "Convolutional Neural Networks for Font Classification". In 2017 *14th IAPR International Conference on Document Analysis and Recognition (ICDAR)* (Vol. 1, pp. 985-990). IEEE.
[13] Ustinova, E., Ganin, Y. and Lempitsky, V. (2017). "Multi-Region Bilinear Convolutional Neural Networks for Person Re-Identification". *14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2017.
[14] Zaidah Ibrahim, Nurbaity Sabri and Dino Isa (2018). "Multi-maxpooling Convolutional Neural Network for Medicinal Herb Leaf Recognition". *Proceedings of the 6th IIAE International Conference on Intelligent Systems and Image Processing, Shimane*, Japan, September 2018, pp. 327-331.
[15] Ahmed, A., &Ozeki, T. (2015, October). "Food Image Recognition by Using Bag-of-SURF Features and HOG Features". In *Proceedings of the 3rd International Conference on Human-Agent Interaction* (pp. 179-180). ACM.
[16] Salem, A., &Ozeki, T. (2015, October). "Face Recognition by Using SURF Features with Block-Based Bag of Feature Models". In *Proceedings of the 3rd International Conference on Human-Agent Interaction*, pp. 309-310. ACM.
[17] Ibrahim, Z., Isa, D., Rajkumar, R., & Kendall, G. (2009, August). "Document Zone Content Classification for Technical Document Images using Artificial Neural Networks and Support Vector Machines. In *Applications of Digital Information and Web Technologies*, 2009. ICADIWT'09. Second International Conference on the, pp. 345-350.
[18] Zaidah Ibrahim, Nurbaity Sabri and Dino Isa (2018). "Leaf Recognition using Texture Features for Herbal Plant Identification". *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)* Vol. 9, No. 1, January 2018, pp. 152~156.
[19] Nurul Fatihah Sahidan, Ahmad Khairi Juha and Zaidah Ibrahim. "Evaluation of Basic Convolutional Neural Network and Bag of Features for Leaf Recognition (2019)". *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)* Vol. 14, No. 1, April 2019, pp. 327-332.
[20] Nik Noor Akmal abdul Hamid, Rabiatul Adawiya Razali and Zaidah Ibrahim (2019). "Comparing Bags of Features, Conventional Neural Network and AlexNet for Fruit Recognition". *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)* Vol. 14, No. 1, April 2019, pp. 333-339.
[21] The Chars74K dataset. Retrieved from http://www.ee.surrey.ac.uk/CVSSP/demos/chars74k/
[22] Gerber, C., & Chung, M. (2016). "Number Plate Detection with a Multi-Convolutional Neural Network Approach with Optical Character Recognition for Mobile Devices". *Journal of Information Processing Systems*, 12(1).
[23] Loussaief, S., &Abdelkrim, A. (2018, March). "Deep Learning Vs Bag of Features in Machine Learning for Image Classification". In *2018 International Conference on Advanced Systems and Electric Technologies (IC_ASET)*, pp. 6-10.
[24] Shima, Y., Nakashima, Y., & Yasuda, M. (2018, April). "Handwritten Digits Recognition by Using CNN Alex-Net Pre-trained for Large-scale Object Image Dataset". In *Proceedings of the 3rd International Conference on Multimedia Systems and Signal Processing*, pp. 36-40. ACM.
[25] Du, G., Su, F., & Cai, A. (2009, October). "Face recognition using SURF features. In MIPPR 2009: Pattern Recognition and Computer Vision (Vol. 7496, p. 749628)". *International Society for Optics and Photonics*.

BIOGRAPHIES OF AUTHORS

Nasibah Husna Mohd Kadir is currently taking Masters Degree in Computer Science (Web Technology) at the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia. Her area of interest includes image processing and artificial intelligence.

Sharifah Nur Syafiqah Mohd Nur Hidayah is currently taking Masters Degree in Computer Science (Web Technology) at the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia. Her current of interest includes image processing and artificial intelligence.

Norasiah Mohammad is a Senior Lecturer at the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia. She has been teaching courses related to Algorithms and Computer Science Education for more than ten years. Currently, she is actively involved in Interactive and Communication Technology (ICCT) research interest group where she has produced a few publications in the related area.

Zaidah Ibrahim is an Assoc. Prof. at the Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia. She has been teaching courses related to Artificial Intelligence for more than ten years. Currently, she is actively involved in Digital Image, Audio and Signal Technology (DIAST) research interest group where she has published some publications in journals and presented at conferences, nationally and internationally.