Improved Handwriting Recognition System using Capsule Network

Tan Jun Yin[*a], Wan Nor Al-Ashekin Wan Husin[b]

[a,b]*Faculty of Engineering, Science and Technology, Nilai University, Malaysia*

No 1, Persiaran Universiti, Putra Nilai, 71800 Nilai, Negeri Sembilan

Abstract

Malay handwriting recognition is playing an important role and very useful in the field of automated cheque processing, automated detection of vehicle number plate and digitization of Malay documents such as textbooks, rare books, newspapers, passports, certificates, manuscripts and others. With the rapid development of artificial and deep learning, many industry has increased their demand for task automation to improve work productivity. In recent years, various type of techniques have been used to perform the image and text detection easier, but there are still challenges and issues with accuracy, performance and reliability. According to the research, the existing system is using the Artificial Neutral Network (ANN) method to recognize the handwriting character and the average accuracy result is below than 90% which is still not accurate enough. The goal of this project is to use deep learning method to increase the accuracy of Malay handwriting recognition. Thus, three deep learning approaches (Convolutional Neutral Networks, Fully Convolutional Network and Capsule Network) will be analyzed in this paper. Since the Capsule Network has an advantage in image and text detection, it has been proved to be able to generate more than 95% of the expected accurate result. Therefore, Capsule Network has been proposed to be applied in the Malay handwriting recognition system to ensure that the recognition system provides high accuracy and satisfaction to the user.

Keywords: Handwriting recognition system, Deep learning, Capsule Network, Capsnet

1. INTRODUCTION

Handwriting recognition system or Optical Character Recognition (OCR) is playing important role for allowing the people to convert the physical documents into machine readable format for further used in data processing purpose. In fact, many organizations are having very large amount of documentation or forms to record the daily data, issues or solution to improve the operation, management and sales performance among many other reasons. OCR technology allows the organizations or people to extract the data automatically from a scanned physical document such as book, receipt, invoice, identify card, license and many more. As the trend of the world is moving toward digitization and big data, the companies need to adopt innovative solutions in their work operation that could reduce the manpower effort, speed up the workflow to meet with the customer needs. It is capable of reducing the total time it takes in human manually data extraction and entry. With high accuracy and robust OCR technology, people can extract the data from multiple electronic or paper document formats. Indeed, traditional OCR achieved very excellent result as high as 99% accuracy (Greg Council, 2018) with the structured printed text and high quality image given but actually it is still difficult to handle the handwritten text due to the reason of irregularities of handwriting, different styles of people writing, inconsistent alignment and spacing between each text and paragraph. Current technology is still very challenging in handwriting recognition because they are unable to recognize the handwritten from documents accurately as they are not designed for these problem and challenges. Due to this reason, recently trend is shifting away to propose better Artificial Intelligent and deep learning algorithms to solve the handwriting issues.

2. PROJECT OBJECTIVE

The proposed objective of this project is to analyze and improve the accuracy in order to develop a more robustness handwriting recognition system by using the proposed technique.

3. RESEARCH BACKGROUND

3.1. Problem Statement

In many organizations, all handwritten information on the physical form such as bills, invoices, receipts and daily record must be transferred to centralized digital database to

ensure the availability. In fact, most organizations currently still using human inspection to manually classify and extract the data from physical documents, which make the process become time-consuming and costly. Therefore, handwriting recognition system can be employed by the organization to digitize them into database. Although the system able to read and extract the document data very efficiently and quickly but the accuracy still a biggest challenge. Indeed, the current handwriting recognition system is still using the traditional approach, namely "Artificial Neutral Network (ANN)". This algorithm work by classifying the text image pattern by learning the pattern and predict based on the structure of neurons. According to the recent experimental result (Mudunuri Prashanth Varma, Shubhro Jyoti Hore, Uday.C, S.Omnath Reddy and Vinay Jha Pillai, 2020), the ANN model only achieved in 88.46% of accuracy rate, which is still not accurate enough. A missing or an error in a documents can greatly impact a decision made by the organization. To prevent these, the workers are still required to use the human view to manually correct the remaining errors.

3.2. Scope

The proposed system should be able to recognize the handwritten character of Malay language, also including both uppercase and lowercase. Following are the scope defined that the proposed system must be able to:

- Allow the system to take handwritten character as the input.
- Allow the system to convert the handwritten character to machine-readable text data as the output.

4. Literature Review

4.1. Convolutional Neutral Network

Convolutional neutral network, usually called CNN that is so far been most commonly used method in the part of analyzing and processing image because it can learn highly abstract features and very efficiently to identify object (Zhang, 2016), The CNN is using the weight sharing concept to reduce the total number of parameter that supposed to be trained, which make the improvement of generalization. Due to the less number of parameters, CNN able to perform training effectively and smoothly in order to generate the output. The model starts with the low-level features extraction such as lines, corners

and edges, then performs to the high-level features recognition and finally the classification of the entire detected elements or object from the input image. Generally, almost all the CNN architectures follow the same working principles which containing of three layers and each of them have performs different functions. The proposed architecture of CNN model consists of three layers: (i) Convolutional layers; (ii) Pooling/Subsampling layers; (iii) Fully connected layers as shown in figure 1.

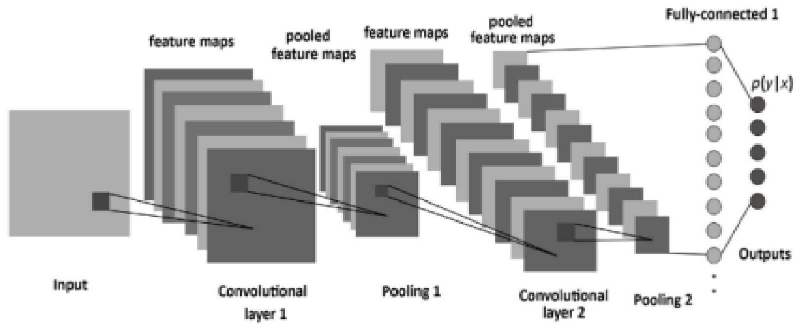

Figure 1 Architecture of CNN (Saleh Albelwi, 2017)

CNN apply a filter to the input image to create a feature map to allow the convolutional layers are able to extract them. The extracted features will be summarized by the pooling layers to reduce the size for training purpose and the fully connected layer for classification are determined. CNN with the help of the collection of training data, the model is allowed be easier in the handwritten characters recognition. CNN allows the handwriting recognition system to be more sensitive to different characteristics of the object. Therefore, it is able easy to classify and recognize different styles of handwriting characters with a higher accuracy result based on the training data and samples.

4.2. Fully Convolutional Network

Since the CNN produce result only for the fixed size inputs that comes from the fully connected layers. To overcome this limitation and make the improvement of CNN's accuracy, Jonathan Longden has introduced the fully convolutional network (FCN) which is trained end to end model for image segmentation (Jonathan Long, Evan Shelhamer and Trevor Darrell, 2015). The model proposed recovers the category of each pixel from the abstract features. In other word, the semantic segmentation problem have been solved by extending the classification from image-level to pixel level. The authors of FCN modifies the well-known classification network AlexNet, VGG net and GoogleNet to enable the model able to detect a flexible fixed size of the input. Unlike CNN model, the fully connected layers are applied after the convolutional layers to generate the features vectors in the fixed length, these layers are declared as convolutions with kernels that covering the entire input region. By casting them into FCN, the model accept the input without any fixed size of images. The workflow of convolutional layers in FCN is illustrated in Figure 2.

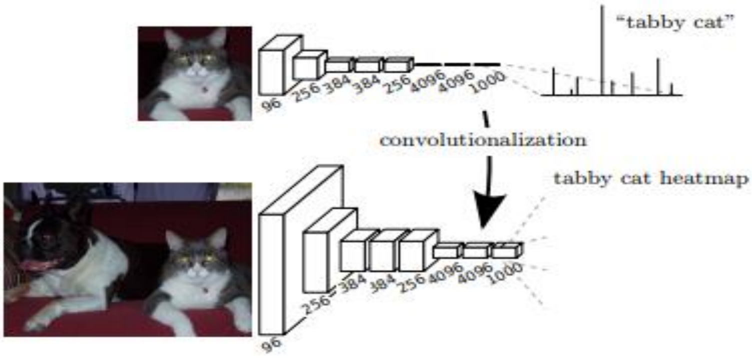

Figure 2 Architecture of FCN (Sik-Ho Tsang, 2018)

FCN able to produce great performance in image recognition and it is an improved version based on CNN model. According to a research report (Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, Alan L. Yuille), researcher only achieved the accuracy result in 79.7% in the collection of the test set. In a word, FCN able to handle the flexible input size of in the input image but the accuracy still not accurate enough even the convolutional layers have replaced by the fully connection layers.

4.3. Capsule Network

To improve the overall performance and overcome the limitation of CNN, a new deep learning architecture, called Capsule Network (CapsNet) has been proposed by Geoffrey Hinton in 2017. Geoffrey and his team also published an new paper to introduce a new algorithm, called "Dynamic routing between capsules" (Sara Sabour,

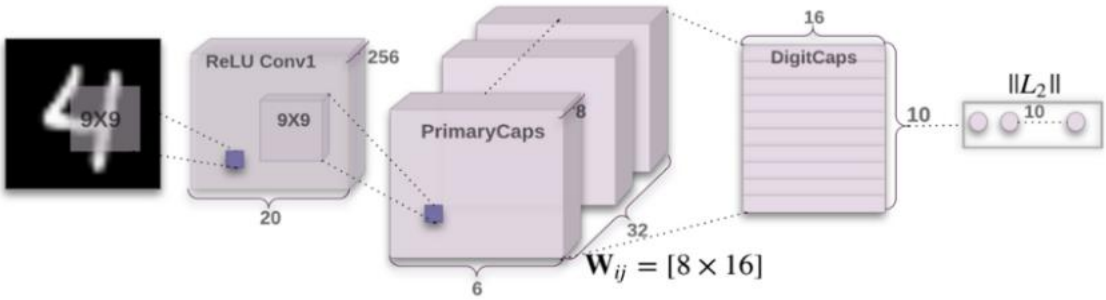

Figure 3 Architecture of CapsNet (Aryan Misra, 2019)

Nicholas Frosst, Geoffrey E Hinton, 2017) .The model contains a set of nested layers, more layers are added inside a single layer. It was completely different from the approach adopted by traditional layered structure of neural networks but solved the several issues faced in CNN. For instance, CNN model requires significant amount of training dataset and sample in the training process. In scenarios where the object recognized is not match with data in the sample dataset and sufficient data for training is not available for training deep neural networks such as CNN model, thus the low accuracy result will be given by the model The next issues is inaccuracy result produced by CNN mode when the object in the image is like insensitivity to feature position and rotation. In 2014, the author of CNN, Geoffrey has admitted his mistakes about the pooling operation (figure 1) that used in CNN to transfer more important features to the next layer. When the object in the image is not aligned or rotated, CNN would not work in invariance by modifying the position of the input to achieve the same output because of max pooling layer in CNN losses valuable information and precise spatial relationship between higher level components. Fortunately, CapsNet have been proved to fix these disadvantages. (Sandeep Pande, Manna Sheela Rani Chetty, 2019). The proposed model consists with the following three layers in architecture of CapsNet: (i) Convolutional layer; (ii) Primary-Capsule layer; (iii) Digit-Capsule layer ; (iv) Fully connected layer, these layers are demonstrated in figure 3.

CapsNet deep learning architecture is robust to affine transformations. Instead of making a neutral network “deeper” in height, it makes it deeper in nesting. According to a research journal (Akmaljon Palvanov and Young Im Cho, 2018), the several deep learning algorithms included CapsNet have been analyzed and final result showed that the very excellent performance result was achieved by the CapsNet. The model produces the highest result among others, also less training data are required to train the model. The only limitation of CapsNet is the model requires the more training time than CNN.

4.4. Dataset - EMNIST

The EMNIST dataset (Gregory Cohen, Saeed Afshar, Jonathan Tapson, André van Schaik, 2017) is an open source dataset for handwritten character classification purpose. It is a set collection of handwritten character and digits that derived from a handprinted and character data “NIST Special Database 19”. The EMNIST dataset contain both 26 classes for uppercase [A-Z] , lowercase [a-z] and 10 classes for digits number [0-9], thus total 62 classes are collected and they will be chosen in this project for training purpose.

Figure 4 Training Samples

4.5. Summary

In a result, the Capsule Network is chosen for handwriting recognition because it show the better performance as compared to CNN and FCN. Since every people has their own different writing style and their written word are not always aligned and consistent as the word is different size, position, rotation, distance and precise, the reason of CNN it is not suitable because of CNN is using the max pooling layer to extract the features but it is not work efficiency like human brain. It must achieve the translation invariance in a much better way but CapsNet do. Besides this, CNN and FCN required heavy amount of training sample, the machine will generate a poor accuracy result if there is no sufficient handwriting training sample, fortunately CapsNet only required the small amount of sample data for training purpose, while still expecting more than 95% of accuracy result, also it can accept any size and position of the input image. The automation with high accuracy and less memory are important issue especially for real-time handwriting recognition applications, thus the CapsNet approach shall be applied to deliver best performance and accurate result in handwriting recognition.

5. RESEARCH METHODOLOGY

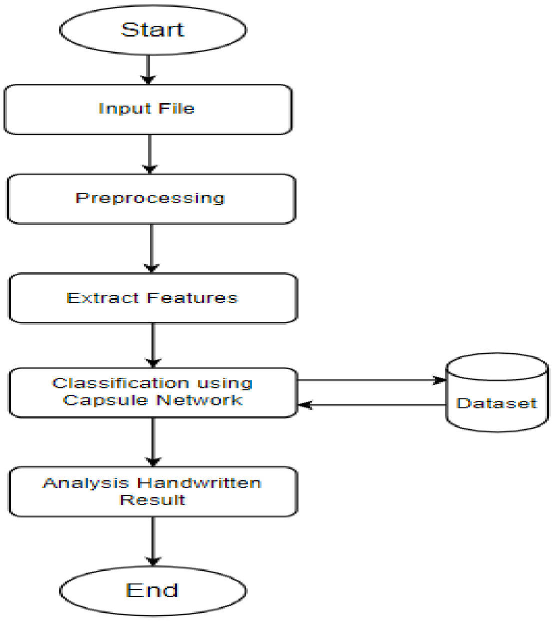

Figure 5 Flowchart of Proposed System

As shown in the flowchart in figure 5, the working flow of the proposed handwriting recognition system begins from the input image file by the user. The system allow to accept any image file in any image format such as "JPEG" or "JPG". To allow the machine more easily to capture the handwritten character in the image file, the preprocessing step is required after the input session. Preprocessing step is process of modify the features in the image to improve the image quality for further action. In this project, RGB converted will be applied to remove the image background pixels and convert the colorful image into black and white, which called as grayscale. Then the thresholding and noise reduction will be implemented to improve the quality and avoid some unnecessary noise element in the image. It will allow the machine becomes more easily to capture the data accurately. After this, the feature extraction step is required to transform the processed image into a collection of the features such as lines, edges, zoning, corner, profiles, crossing and distances from the local text region, thus the machine able to understand and learn the features from the represented input data. After this, the Capsule Network is used as a classifier for classifying handwritten character purpose. The model will extract the combination of the detected features in order to extract and recognize the details features to improve the accuracy result. Moreover, the training phase is required to train the model by supplying the trading data from

EMNIST dataset as mentioned in the section 4.4. As a result and they will result in different classification output, therefore the machine able to recognize the handwritten input data after the analysis and the machine readable format will be display to the user as the final output.

6. RESULTS AND DISCUSSIONS

Methods/Techniques	Input Image	Accuracy
Artificial Neutral Network(ANN)	Handwritten text	88.46%
Capsule Network (CapsNet)	Handwritten text	**95% (Expected)**

Table 1 Comparison of various type of proposed methods

From the table above, the existing and proposed deep learning methodology have been compared by using the handwritten text image. According to the experimental result, the existing method ANN only obtains accuracy of 88.46% in handwriting, and many features from the input handwritten text image were difficult to be extracted and recognized. (Mudunuri Prashanth Varma, Shubhro Jyoti Hore, Uday.C, S.Omnath Reddy and Vinay Jha Pillai, 2020). From what have been analyzed and expected in earlier sessions, the input text image should be pre-processed, features extracted then classified by proposed Capsnet to extract more detailed features and improve the accuracy result, thus the proposed CapsNet methodology is expected to achieve 95% of accuracy result for the handwriting recognition to provide more robust handwriting recognition system.

7. CONCLUSION

In conclusion, this project is to build the handwriting recognition system to recognize the handwritten of Malay character and generate the machine-readable format as the final output. In order to deliver the higher accuracy to the user, Capsule Network deep learning method is proposed after doing the research and comparison with other techniques. There are some recommendations could be explored, the system should be implemented for supporting more than one language such as Chinese and Tamil into our handwriting recognition system. In addition, the system also can provide the translation feature to allow the translate the document to another chosen language. To conclude in fact, although the Capsule Network still new in the deep learning area but it was achieved very excellent result as it has many advantages in the part of image and text classification among many other traditional approaches.

8. REFERENCES

Akmaljon Palvanov and Young Im Cho. (2018). Comparisons of Deep Learning Algorithms for MNIST in Real-Time Environment. *International Journal of Fuzzy Logic and Intelligent Systems, 18*(2), 126-134.

Gregory Cohen, Saeed Afshar, Jonathan Tapson, André van Schaik. (2017). EMNIST: an extension of MNIST to handwritten letters.

Jiuxiang Gua,Zhenhua Wang,Jason Kuen. (2017). Recent Advances in Convolutional Neural Networks. 28.

Jonathan Long, Evan Shelhamer and Trevor Darrell. (2015). *Fully Convolutional Networks for Semantic Segmentation.* IEEE Transactions on Pattern Analysis and Machine Intelligence.

Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, Alan L. Yuille. (n.d.). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs.

Mudunuri Prashanth Varma, Shubhro Jyoti Hore, Uday.C, S.Omnath Reddy and Vinay Jha Pillai. (2020). Optimized Handwritten Character Recognition Using Artificial Neutral Network. *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH, 9*(1), 4442-4445.

Fadzil, P. Z. (2017). *Medical Records Contents Scanning*. Retrieved 3 12, 2020, from http://www.myhealth.gov.my/en/medical-records-contents-scanning/

Greg Council. (2018). Using OCR: How Accurate is Your Data? Retrieved from https://tdwi.org/articles/2018/03/05/diq-all-how-accurate-is-your-data.aspx

Riter. (2018). Capsule Networks As a New Approach to Image Recognition. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6108980/

Sandeep Pande, Manna Sheela Rani Chetty. (2019). Analysis Of Capsule Network (Capsnet) Architectures And Applications. *Journal of Advanced Research in Dynamical and Control Systems*, *10*(10).

Sara Sabour, Nicholas Frosst, Geoffrey E Hinton. (2017). Dynamic Routing Between Capsules. *Computer Vision and Pattern Recognition*, 1-11.

Zhang, Z. (2016). *Derivation of Backpropagation in Convolutional Neural Network (CNN).* Hairong Qi.

Aryan Misra. (2019). *Capsule Networks: The New Deep Learning Network*. Retrieved from https://towardsdatascience.com/capsule-networks-the-new-deep-learning-network-bd917e6818e8

Mudunuri Prashanth Varma, Shubhro Jyoti Hore, Uday.C, S.Omnath Reddy and Vinay Jha Pillai. (2020). Optimized Handwritten Character Recognition Using Artificial Neutral Network. *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH, 9*(1), 4442-4445.

Saleh Albelwi. (2017). A Framework for Designing the Architectures of Deep Convolutional Neural Networks.

Sik-Ho Tsang. (2018). *Review: FCN — Fully Convolutional Network (Semantic Segmentation)*. Retrieved from https://towardsdatascience.com/review-fcn-semantic-segmentation-eb8c9b50d2d1