Inception Network for Baybáyin Handwriting Recognition

James Arnold E. Nogra

Cebu Institute of Technology – University, jamesnogra@gmail.com

ABSTRACT

When the Philippine government legislated that the Baybáyin should become the Philippines ' national writing system, educational institutions and other cultural organizations started planning to incorporate this writing method into the existing educational curriculum. One way back to implement this writing method is by using a smartphone app. As of October 2018, an eLearning app named Learn Baybáyin has been in the Android Playstore. This eLearning app will teach anyone how to read and write the script for Baybáyin. This mobile app is powered by a convolutional neural network (CNN) and was trained with 94% accuracy on the validation data. This study aims to replace the existing CNN model with an inception network because this type of neural network can successfully classify patterns or details regardless of its size in an image. After testing an inception model five times, the average validation accuracy was 96.2% which is higher than the current CNN model. The inception network got a higher validation accuracy than the previous LSTM (92.9%) and CNN (94%) model. Because of this, the inception network will replace the current model used by the eLearning app Learn Baybáyin to recognize handwritten characters.

Key words: CNN, Handwriting Recognition, Inception Network, LSTM, Optical Character Recognition

1. INTRODUCTION

1.1 Background of the Study

House Bill 1022 states that the Baybáyin is known as the Philippines ' national writing. The house bill was made into effect last April 2018. Since then, many cultural and educational institutions such as the National Commission for Culture and Arts and the Department of Education showed their support in terms of the implementation. Some ancient artifacts found around the Philippines like vases and tablets have these writing systems on it. This script is considered as one of the pre-colonial writing systems of the Philippines. Even though this writing system is used whole throughout the country, evidence suggests that this has been more widely used in the northern part of the Philippines specifically the Tagalog people.

The Baybáyin writing system has three vowels and fourteen consonants. The three vowels are A, E or I, O or U. E and I and O and U have the same character representation. The fourteen consonants are B, K, D/R, G, H, L, M, N, NG, P, S, T, W, and Y. In some variations, D and R have different characters but in this study, it's only represented by one. In total, there are 59 classes of Baybáyin characters in which 56 are comprised of the variations of the consonant characters.

Numerous studies are currently out there employing handwriting recognition. Usually, these optical handwriting recognition studies are based on the application of more than one algorithm. A unique implementation for handwriting recognition for this writing system would require an Inception Network. This type of neural network will allow for image recognition where the sample sizes have different sizes [1]. Convolutional neural networks (CNN) and long short-term memory (LSTM) have also been used in image recognition or classification. These two neural networks have proven to have a high accuracy rate in terms of image classification. But in terms of the training data, these two neural networks need a lot of sample images. The problem arises when there are only a few hundred sample images with different sizes. An inception network can solve this issue because it processes an image through a series of convolutions with different filter sizes [2].

1.2 Objectives of the Study

One of the aims of this study is to compare whether an Inception network would have a better validation accuracy than the current CNN and LSTM models. A comparison between the CNN, LSTM, and the proposed inception network must be made in order to assess which one is better. If the inception network will be better than the existing LSTM and CNN models in terms of validation accuracy, then it will replace the backend of the eLearning app Learn Baybáyin.

1.3 Scope and Limitation

The inception neural network that will be developed will only be able to classify the Baybáyin characters one at a time. A word or sentence recognition will not be part of this study. The image being classified must also have a white background and the color of the handwriting must be black.

2. REVIEW OF RELATED LITERATURE

The first mention of the Baybáyin writing system is in the text Doctrina Christiana which was written in 1593. This was a text written by a Spanish priest named Juan de Plasencia and is considered to be one of the oldest books of the Philippines. The text mentioned that this writing system has been widely used by the Filipinos before they arrived. The arrival of the Spaniards in the Philippines also marks the start of the death of this writing system. Currently, there are organizations and groups that aim to revive this writing system. They usually give talks and short courses on this writing system to grade school and high school students.

Studies like character recognition for Baybáyin handwriting already exists and it has been used in eLearning applications. Convolutional neural networks (CNN) have been used to classify Baybáyin characters. LSTM networks have also been used to perform the same task [3]. The CNN model has achieved 94% accuracy using the validation data while the LSTM model obtained 92.9% validation accuracy. Clearly, CNN outperforms the LSTM model in the Baybáyin handwriting recognition but these two models are trained using sample images that are drawn in almost exactly the same size. CNN has been proven to be far more superior to almost any other neural networks in terms of image classification [4]. Other optical character recognition (OCR) techniques are also used to classify handwritten digits or characters. One of the OCR techniques used for character recognition Support vector machines (SVM) is one of the OCR techniques used for character recognition. This algorithm is used to identify hand-written Telugu characters [5]. These are other algorithms or neural networks for optical character recognition but the most widely used are convolutional neural networks.

To solve the issues in training a neural network model that has different training subject sizes in the image, an inception network must be used. This type of neural network can identify details that appear at different sizes in different images [6]. For instance, one image might have an eye that covers the entire resolution of the image while another will have the face of a person that has both eyes. Both of these images have eyes on it but in different sizes. Similarly, a person might draw a smaller character on a canvas while another draws the same character on canvas but in a different size. These two characters must be classified as the same characters. Inception networks have also been used to classify different flower species [7]. This type of neural network model has even been used in medical applications such as pathogen detection which got high accuracy [8]. Inception networks have many applications in the real world, especially in image classification or recognition.

3. METHODOLOGY

Since October 2018, an established eLearning app called Learn Baybáyin has been in the Android PlayStore. This smartphone app featured in Figure 1 Teaches its users how to read and write the Baybáyin script. A convolutionary

neural network (CNN) is the model used by this mobile application for classifying user's handwritten characters. When trained, this CNN model achieved a validation accuracy of 94%. Another model, the long short-term memory (LSTM) model was also tested but the model only got a validation accuracy of 92.9%. Even though the CNN model has a validation accuracy high enough for the classification task, inception networks are known to have better accuracy.
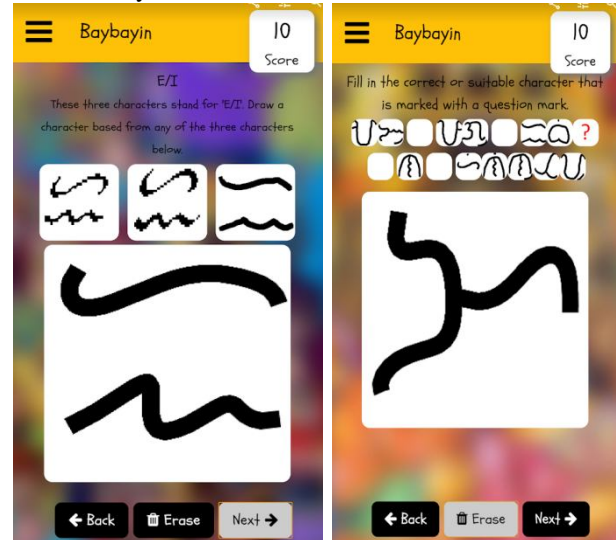

Figure 1: Sample screenshots of the eLearning app Learn Baybáyin.

The handwritten drawn characters Baybáyin that was used in this study were obtained from the previous studies. The hand-drawn characters were made by 25 people ranging from the ages of 10 to 55. These drawn Baybáyin characters are then manually checked if it is accepted or not. Hand-drawn characters that are not acceptable are deleted. A mobile app is used to let users draw the sample handwritten images. In the mobile app, the contributors are presented with three sample characters where they create their own in a canvas shown in Figure 2.
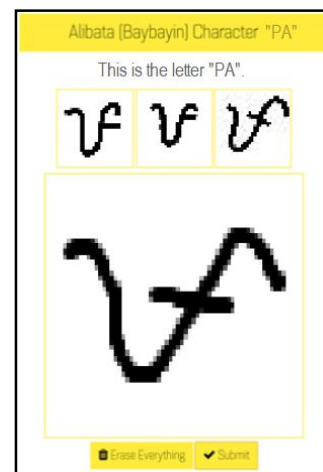

Figure 2: The mobile app where contributors can draw their own version of a Baybáyin character based on the three sample characters presented.

For this study, there were about 12,545 images gathered. Each class or character has at least 200 sample drawn characters. In the previous studies, the LSTM and CNN models only used less than 10,000 images during the training of both models. All of the images gathered are resized to 28×28 pixels and converted to a binary image, all the pixels are either black or white.
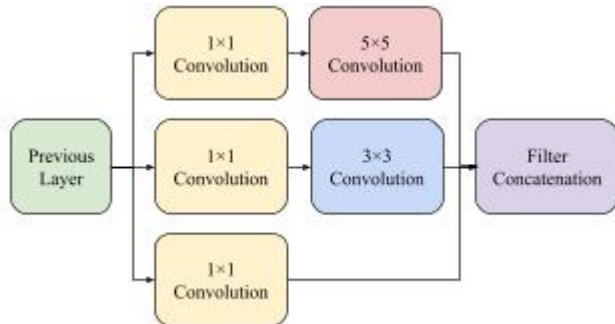

Figure 3: Diagram of the inception module used in handwriting recognition.

In each of the inception modules, 1×1 convolutions are applied to reduce the number of feature maps as shown in Figure 3. The startup module implements a reduction of the dimensions to minimize the number of function maps[9]. The less function maps a startup module has, the less time it would require for preparation.

Figure 4: The complete structure of the inception network used in Baybáyin character recognition.

For the inception network, a total of three inception modules are connected shown in Figure 4. All of the three inception modules are identical and then followed by a max pooling layer. After three three inception modules is a fully connected layer with 1×1×256 dimensions. A dropout layer of 80% is applied after the first fully connected layer. This means that only 80% of the units are retained during training the model. The final layer of the inception network is a fully connected layer with 1×1×59 dimensions to represent the number of classes in the Baybáyin handwriting system. The final layer also uses the Softmax as an activation function. This Softmax function calculates the probability distribution for all classes [10]. The convolution and the first fully connected layer uses the rectified linear unit (ReLU) activation function. This activation function returns either a positive number or zero. Many neural networks use this activation function because it is easier to train and most of the time achieves better performance [11].

80 per cent of the sample images are used to train the model during the model testing. The other 20 per cent is used to validate the model being educated. Up to 50 epochs were equipped for the inception model. To keep track of the training phase, the training loss, the validation loss, training accuracy and validation accuracy are registered. For training the model, a computer with a 4-core Intel i7-7700HQ processor with 16 gigabytes of memory was used.

4. RESULTS AND DISCUSSION

Training the model with more than 12,000 training images at 28×28 resolution took almost three hours. The initiation model was trained 5 times and the cumulative loss of testing, loss of validation, accuracy of testing, and accuracy of validation was tabulated. The loss of training and validation was reported to test that the model did not over fit [12]. The research methods of this study have been influenced by [13], [14], and [15].

Table 1: Table of the average training loss and validation loss of the five training tests done by the inception network for 50 epochs.

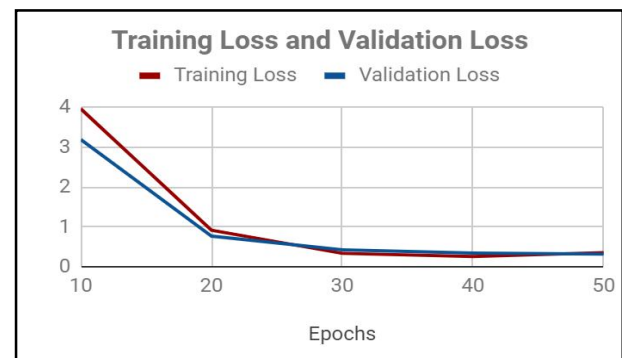| Epochs	Training Loss	Validation Loss
10	3.95	3.18
20	0.92	0.77
30	0.34	0.43
40	0.26	0.35
50	0.36	0.32

Figure 5: Graph of the average training loss and validation loss of the five training tests done by the inception network for 50 epochs.

As shown in Table 1 and Figure 5, the average training loss and validation loss are almost equal which means the model didn't fit. Another evidence that the model didn't fit is that both of the values are close to zero despite the fact that there are 59 classes in this handwriting recognition.

Table 2: Table of the average training accuracy and validation accuracy of the five training tests done by the inception network for 50 epochs.

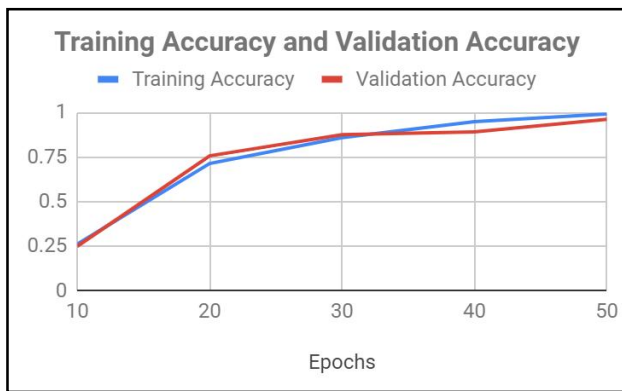| Epochs	Training Accuracy	Validation Accuracy
10	0.261	0.249
20	0.714	0.757
30	0.859	0.875
40	0.948	0.891
50	0.992	0.962

Figure 6: Graph of the average training accuracy and validation accuracy of the five training tests done by the inception network for 50 epochs.

After the five tests with 50 epochs each, the average training accuracy of the inception model is 0.992 while the average validation accuracy is 0.962 as shown in Table 2 and Figure 6. This validation accuracy is high enough to be used as the backend of an eLearning app that will classify Baybáyin handwritten characters. As shown in Figure 7, it can classify handwritten training images with very high accuracy.
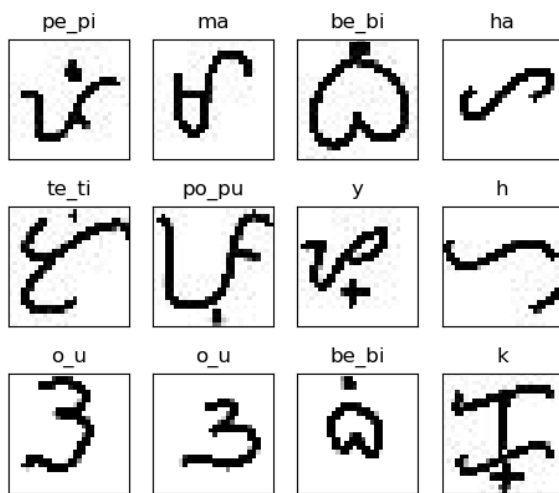

Figure 7: Sample training output of the inception network after 50 epochs where all 12 random characters from the validation data are classified correctly.
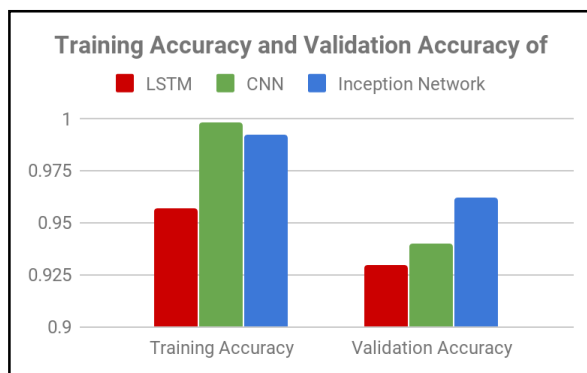

Figure 8: Graph of the comparisons of the training and validation accuracies of the CNN, LSTM, and the new inception network for Baybáyin handwriting recognition.

Based on Figure 8, the validation accuracy of the previous LSTM model is 0.929, the CNN model has 0.94, while the new inception network has 0.962. This makes the inception model superior to the two previous neural networks used to perform the Baybáyin handwriting recognition. Even though the inception network has a higher validation accuracy, it is also worth noting that when this model was trained, it has more than training images than the previous LSTM and CNN studies.

5. CONCLUSION

After training an inception model five times, it can be inferred that the average training and validation accuracy of this model compared to the existing LSTM and CNN models are better. The inception model got an average validation accuracy of 96.2% compared to 94% of the CNN model and 92.9% of the LSTM model. The inception model tested has three starting modules followed by a fully connected layer, then an 80 percent dropout, and then a full connected output layer. This model is a viable candidate for replacing the backend of the eLearning app Learn Baybáyin to identify Baybáyin handwriting characters due to the high precision of this neural network. The model used in this study might also be used to detect other variations of the Baybáyin handwriting system.

REFERENCES

1. Z. Xiong, Y. Yuan and Q. Wang, **AI-NET: Attention Inception Neural Networks for Hyperspectral Image Classification**, *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, Valencia, 2018, pp. 2647-2650. https://doi.org/10.1109/IGARSS.2018.8517365

2. C. Chen and F. Qi, **Single Image Super-Resolution Using Deep CNN with Dense Skip Connections and Inception-ResNet**, *2018 9th International Conference on Information Technology in Medicine and Education (ITME)*, Hangzhou, 2018, pp. 999-1003. https://doi.org/10.1109/ITME.2018.00222

3. J. A. Nogra, C. L. S. Romana and E. Maravillas, **LSTM Neural Networks for Baybáyin Handwriting Recognition,** *2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS)*, Singapore, 2019, pp. 62-66.

4. D. Ciresan, U. Meier, J. Masci, L. Gambardella, J. Schmidhuber, **Flexible, High Performance Convolutional Neural Networks for Image Classification**, *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, Retrieved from https://www.aaai.org/ocs/index.php/IJCAI/IJCAI11/paper/view/3098/3425

5. N. Prameela, P. Anjusha and R. Karthik, **Off-line Telugu handwritten characters recognition using optical character recognition**, *2017 International conference of Electronics, Communication and*

Aerospace Technology (ICECA), Coimbatore, 2017, pp. 223-226.
https://doi.org/10.1109/ICECA.2017.8212801

6. C. Szegedy, S. Ioffe, and V. Vanhoucke, **Inception-v4, inception-resnet and the impact of residual connections on learning**, *CoRR*, abs/1602.07261, 2016.

7. Xiaoling Xia, Cui Xu and Bing Nan, *Inception-v3 for flower classification*, *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, Chengdu, 2017, pp. 783-787.

8. R. Sinha and J. Clarke, **When technology meets technology: Retrained 'Inception V3' classifier for NGS based pathogen detection**, *2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Kansas City, MO, 2017, pp. 1-5.

9. C. Szegedy et al., **Going deeper with convolutions**, *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 2015, pp. 1-9.
https://doi.org/10.1109/CVPR.2015.7298594

10. C. Nwankpa et al., **Activation Functions: Comparison of Trends in Practice and Research for Deep Learning**, arXiv:1811.03378v1 [cs.LG] 8, November 2018

11. Z. Hu, Y. Li and Z. Yang, **Improving Convolutional Neural Network Using Pseudo Derivative ReLU**, *2018 5th International Conference on Systems and Informatics (ICSAI)*, Nanjing, 2018, pp. 283-287.

12. I. Bilbao and J. Bilbao, **Overfitting problem and the over-training in the era of data: Particularly for Artificial Neural Networks**, *2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)*, Cairo, 2017, pp. 173-177.

13. G. Alcober, T. Revano, & M. Garcia, **E-Safety in the Use of Social Networking Application**, *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 9, No. 1.2 2020,
https://doi.org/10.30534/ijatcse/2020/1291.22020

14. R. Dellosa, **An Efficient Position Estimation of Indoor Positioning System Based on Dynamic Time Warping**, *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 9, No. 1.2 2020, https://doi.org/10.30534/ijatcse/2020/0491.22020

15. J. Victoriano & L. Lacatan, **A Geospatial Analysis and Kernel Density Estimation of River Quality Parameter in Bulacan, Philippines**, *International Journal of Advanced Trends in Computer Science and Engineering*, Vol. 9, No. 1.2 2020,
https://doi.org/10.30534/ijatcse/2020/1191.22020