Comparison of Activation Function for Offline Handwritten Kanji Document Detection using Convolutional Neural Network

Anthony Adole[1], Eran Edirisinghe[1], Baihua Li[1], Chris Bearchell[2]

[1] Loughborough University

Loughborough, Leicestershire, LE11 3TU, UK

[2] Surface Intelligence Ltd.

Unipart House, Garsington Rd, Oxford, OX4 2PG, UK

ABSTRACT

In an offline kanji handwriting detection and recognition system, the ability of the neural network to correctly recognise each handwritten character within a document tends to be a significant problem. However, the present state-of-the-art neural network adopted for the object detection task settle for the object location principle but cannot achieve complete detection and lacks the proper use of an activation function. Also, there appears to be a lack of research focusing on developing an activation function that can perfectly enhance the learning ability of an artificial neuron used in a deep neural network model. Therefore, this research paper presents a visual evaluation between monotonic and non-monotonic activation function

Human Interaction & Emerging Technologies (IHIET-AI 2022)
Artificial Intelligence & Future Applications
https://openaccess.cms-conferences.org/#/publications/book/978-1-7923-8989-4

performance effect on a neural network. The results obtained show that the non-monotonic activation functions outperformed the monotonic activation function by achieving a fast speed for detection and recognition of the kanji handwritten characters.

Keywords: Kanji handwriting document, Activation function, Convolutional Neural Network

INTRODUCTION

It has been noted from previous research by scholars from the computer science field that researchers in the neural network have long investigated biological science for inspiration towards the development of neural network models (Cox & Dean 2014). Therefore, the idea of perceptron came into existence due to its description as a single neuron model, which is an antecedent to a more extensive neural network. For knowledge acquisition with a neural network, specific activation functions are implemented within each layer. The non-linearity of a neural network is obtained because of the activation function when the weights of the inputs are summed and passed out of the neuron. The presence of an activation function in a neural network enables deep neural networks to learn a complex task. However, most of the existing activation functions are differential and continuous apart from the rectified unit at '0' (Shanmugamani 2018). The activation function would appear differential when its derivative appears at all points within the domain. It would appear continuous when every little change in input would create a slight change in the output. Therefore, to have a robust neural network that can learn non-exclusive classes (i.e. multi-label classification) and correctly recognise and detect the actual object type, the proper activation function should be implemented within the neural network's hidden and feature extraction layers.

However, the offline handwriting kanji document dataset was used to achieve this research towards multi-classification of objects not minding the location. Handwriting is a skill that everyone learns from a tender age and has distinct fundamental characteristics. It consists of an artificial graphical mark written on a surface used for communication. It serves as a medium that helps transfer information; thereby, the purpose is achieved by the graphical marks that are conventionally related to a language (Plamondon & Srihari 2000). The act of writing has been considered to have made communication and historical transfer of knowledge possible between cultures and civilisations (Plamondon & Srihari 2000). Each script has symbols known as characters of letters, with specific basic and different shapes for identification. There are rules for combining letters to represent higher-level linguistic units. Handwriting character recognition is a technology widely used in the modern world, but it still has some critical challenges. In recent years, handwriting recognition has been one of the foremost fascinating and complicated research areas within image processing and pattern recognition. Therefore, this research aims to achieve the complete cognitive ability of a deep neural network model for offline kanji handwriting documents. To this end, the selected neural network model must detect, recognise, and contextualise handwriting

object types on an offline kanji document to achieve a complete cognitive performance of the deep neural network model. This paper investigates the use of non-monotonic activation function (HardSwish, Mish) and monotonic activation function (Leaky-Relu) for offline multiclassification handwritten document task using visual evaluation for speed and number of characters detected and recognized using each activation function. The investigation was done with the use of YOLOv5 deep neural network model to compare the performance of these activation function concern which is best to be used for vision related task.

RELATED WORKS

After 1951, more inventions have appeared with improved algorithms for optical character recognition; but problems still exist with unusual characters set, font and documents of poor quality. Researchers like Graves et al, 2009 (Graves & Schmidhuber 2009), combined two recent innovation from the ideology of the neural network towards achieving the aim of their research. They combined multidimensional recurrent neural network and connectionist temporal classification to introduce a globally trained offline handwriting recognizer that takes raw pixel data as input. Unlike the competing systems, it does not require any alphabet specific pre-processing stage and can be used for any language. The evidence of its generality and power is provided by the data obtained from an Arabic recognition competition, where it outperformed all entries even though neither of the authors understood Arabic (Graves & Schmidhuber 2009). Another amazing research work that was carried out in this field is the work by Yuan et al, 2012 (Yuan et al. 2012). Aiquan et al made use of a modified LeNet-5 convolutional neural network which they implemented on an offline handwritten English character recognition system. The system had a unique setting for the number of neurons in each layer and ways in which some layers are connected. The output of the convolutional neural network are then set with error correcting codes; thus the convolutional neural networks can reject recognized results. For training of the convolutional neural network, they developed an error sample-based reinforcement learning strategy. However, the experiments are evaluated using the UNIPEN lowercase and uppercase datasets to achieve a recognition rate of 93.7 percent for uppercase and 90.2 percent for lowercase respectively(Yuan et al. 2012). The research carried out by Wu et al, 2014 focused on the convolutional layer of a CNN network architecture. They informed that relaxation convolution layer adopted in their R-CNN does not require neurons within a feature map to share the same convolutional kernel. Therefore, unlike the traditional convolutional layer, endowing the neural network with more power would increase the model's accuracy. When relaxation convolution sharply increases the total number of parameters in the proposed network structure, the adopted alternate training in ATR-CNN regularizes the neural network during the training procedure. They informed that their previous CNN took first place in the Chinese handwriting character recognition competition. At the same time the ATR-CNN which is the present network developed by his team, outperformed their previous one and achieves

state-of-the-art accuracy with an error rate of 3.94 percent, further narrowing the gap between machine and human observers (Wu et al. 2014). A deep convolutional neural network method for HCCR was proposed by Zhuo et al 2015. Their proposed method uses four inception modules to construct an efficient deep network (Zhong et al. 2015). They adopted three types of directional feature map which are gabor, gradient, and HoG feature maps. This directional feature map were used to enhance the performance of GoogleLeNet. They evaluated their proposed network with the ICDAR 13' offline HCCR dataset. The obersavtion informed their network was superior in terms of both accuracy and storage performance for single and ensemble CNN models with an error rate of 3.26 percent (Zhong et al. 2015).

METHODOLOGY

The origin of the Yolo neural network model architecture was inspired by GoogleNet (Redmon & Farhadi 2017). Also, the Yolo neural network model has obtained recognition as one of the most outstanding achievements in the field of real-time object detection and recognition using a deep neural network model. From the research publication by Huang et al. 2017 (Huang et al. 2017), the information obtained informed that the Yolo deep neural network model is known to convert object detection task on localization and classification into regression task (Alganci et al. 2020). It was also known to perform prediction for the final output from tensor obtained after down-sampling of the image data. However, it performs the same steps as the region of interest pooling layer of the faster-rcnn model when it obtains tensor from the down-sampling image data. In this research YoloV5 model which is the current version of the Yolo model family was used to evaluate the performance of the selected activation function. The architecture for the YoloV5 neural network as shown in Figure 1, consists of three essential parts like its predecessors. These parts are model head, model neck, and model backbone. The model backbone adopts the bottleneck cross-stage partial network which enhance and enrich the features extracted form an input data, while the model neck adopts a Path aggregation network which helps with object scaling. The model head used in this architecture is the same as the YoloV4 whereby the primary purpose is to perform the final detection using anchor box.

ACTIVATION FUNCTION SELECTION

In a deep neural network, when obtaining the output of a node, a thing function is used thereby describing this function as one of the building blocks in a neural network. This thing function is called an activation function and it can be monotonic and non-monotonic in suppressing irrelevant data information (i.e. adjusting the gradient information). An activation function tends to output a small value for small data input and a large output value when input data exceeds a specific threshold.

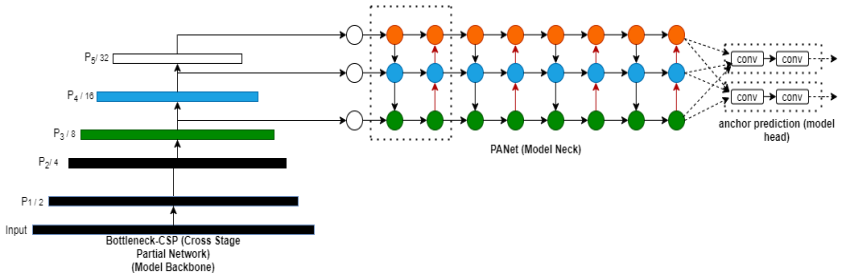

Figure 1. YoloV5 Neural Network Architecture

Therefore, an activation function would only fires when the input data is large just like the action potential life cycle(Molecular Devices 2020). For a neural network to learn a difficult task such as image recognition, non-linearities would be introduced by using an activation function. The output from an activation function of a neuron is passed onto the next layer and the process is repeated all through the layers of the neural network. Therefore, the simple understanding of an activation function effects in neuron informs that it calculates the weighted sum of the input data, adds bias, then decides either to fire or not.

The selection of a good activation function impacts the neural network dynamics both in training and testing (Ramachandran et al. 2017). Currently, the widely-used activation function in the hidden layer of object detection and recognition neural network architecture is the Leaky-Relu which is classified as a monotonic type of an activation function(ProgrammerSought n.d.)(Ultralytics 2020). A monotonic type of activation function is either entirely non-decreasing or non-increasing (Szandała 2021). For the non-monotonic type of activation function, HardSwish and Mish activation function were evaluated visually using the synthetic kanji handwriting dataset. The HardSwish activation function is a variant of the Swish activation function (ProgrammerSought n.d.). The HardSwish has a faster computation power because its algorithm does not have an exponential calculation (Ramachandran et al. 2017) as Swish does. On the other hand Mish activation function has been recorded to outperform Swish activation function (Zhang et al. 2019). The Mish activation function is bounded below and unbounded above same as the HardSwish activation function but at the moment there appears to be no research to evaluate the performance for both type of non-monotonic type of activation function.

RECORD OF EXPERIMENTS

This section presents the dataset generated for the research and results of the experiment performed after adopting either of the activation function into the architecture of the neural network. The adopted techniques requires the use of these activation function independent use of each other in the architecture of the YoloV5 neural network model. This experiment involves the use of three thousand different type of handwritten character to create the synthetic generated kanji handwritten document dataset. In other to have a robust dataset, augmentation steps were adopted

at the pre-processing and post processing stages of the dataset. Figure 2 presents a visual representation of the generated synthetic document used for training the neural network.

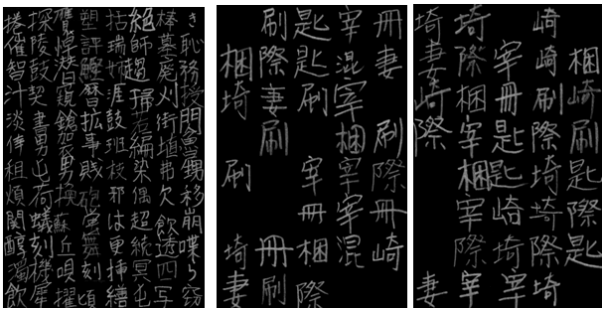

Figure 2. Generated Synthetic Kanji Handwriting Document Dataset

EVALUATION OF TRAINING RESULTS

Before training the YoloV5 neural network model, three case-samples for training were made. Each case-sample adopts the use of either Leaky-Relu, HardSwish, or Mish activation function within the hidden and feature extraction layer of the neural network architecture. Precision and Recall was used as the evaluation metrics for each case-sample during training. The precision metrics would help to determine how accurate the model predictions would be and the recall metrics would help to determine how well the model performs towards detecting all the positive case. Table 1 presents the evaluation for each case-sample using the mean average precision while Table 2 present the evaluation results based on the precision and recall metrics. The mean average precision is computed by taking the average of all the classes (3000 classes) used to training the model.

Table 1: Mean average precision of the activation function obtained during training

Activation Functions	**mAP@0.5**	**mAP@0.5:0.95**
HardSwish	**95.6%**	**93%**
Mish	92.3%	86.3%
Leaky-Relu	86%	76.7%

Table 2: Evaluation of the activation function obtained during training

Activation Functions	**Precision**	**Recall**
HardSwish	75.6%	95.4%
Mish	69.3%	92.2%
Leaky-Relu	76%	81.4%

During the testing stage using the inference graph obtained from case-sample of each

activation function within the architecture of YoloV5, it shows visually base on computation time that the HardSwish outperformed the Mish and Leaky-Relu activation function in performing multiclassification task. Unlike the paper on faster-rcnn (Adole et al. 2020) that uses linear activation for the bound-box regression and was not able to achieve complete detection and recognition. Table 3 shows the inference time of classes correctly detected and recognized. During testing and evaluation the observation made informs that, the inference speed varies due to the computers processing power, the size of the image file, and the number of classes detected and recognized.

Table 3: Inference Time during testing on seven images

Activation Functions	**Inference Time(s) for 7 images**
HardSwish	0.574
Mish	0.523
Leaky-Relu	0.584

CONCLUSIONS

This paper presents the visual performance comparison of the Leaky-Relu, Mish, and HardSwish activation function in a deep neural network architecture to enhance learning a new task. It compares there performance towards multiclassification when more than a thousand classes are involved. Observation from the experiment results presents, the use of a deep neural network model to predict characters on offline handwritten documents for detection and recognition tasks is a means of answering the research question in the field of document recognition and detection. The dataset used for the experiment is a synthetic document generated for the task of multiclassification. It was observed that the choice of activation function within architecture of the feature extraction layer, also has a significant impact on the accuracy and speed of processing. As shown with the aid of the evaluation and testing result presented, the HardSwish activation function improved the detection and recognition ability when tested in real-time compared to the others. Therefore, enabling the achievement of a complete and accurate multiclassification in the field of offline kanji handwriting document detection and recognition.

REFERENCES

Adole, A. et al., 2020. Investigation of Faster-RCNN Inception Resnet V2 on Offline Kanji Handwriting Characters. In *ACM International Conference Proceeding Series*.

Alganci, U., Soydas, M. & Sertel, E., 2020. Comparative research on deep learning approaches for airplane detection from very high-resolution satellite images.

Remote Sensing, 12(3).
Cox, D.D. & Dean, T., 2014. Neural networks and neuroscience-inspired computer vision. *Current Biology*, 24(18), pp.R921–R929.
Graves, A. & Schmidhuber, J., 2009. Offline handwriting recognition with multidimensional recurrent neural networks. *Advances in Neural Information Processing Systems*, 21, pp.1–8. Available at: http://www.idsia.ch/~juergen/nips2009.pdf.
Huang, J. et al., 2017. Speed/accuracy trade-offs for modern convolutional object detectors. In *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. pp. 3296–3305.
Molecular Devices, 2020. What is an action potential. Available at: https://www.moleculardevices.com/applications/patch-clamp-electrophysiology/what-action-potential#gref [Accessed May 8, 2020].
Plamondon, R. & Srihari, S.N., 2000. Online and off-line handwriting recognition: a comprehensive survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), pp.63–84.
ProgrammerSought, Swish & hard-Swish. Available at: https://www.programmersought.com/article/67734330291/ [Accessed December 1, 2020].
Ramachandran, P., Zoph, B. & Le, Q. V., 2017. Searching for activation functions. *arXiv*, pp.1–13.
Redmon, J. & Farhadi, A., 2017. YOLO9000: Better, faster, stronger. In *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. pp. 6517–6525.
Shanmugamani, R., 2018. *Deep Learning for Computer Vision: Expert techniques to train advanced neural networks using TensorFlow and Keras*, Packt Publishing Ltd.
Szandała, T., 2021. Review and comparison of commonly used activation functions for deep neural networks. *Studies in Computational Intelligence*, 903, pp.203–224.
Ultralytics, 2020. Yolov5. Available at: https://github.com/ultralytics/yolov5 [Accessed February 7, 2020].
Wu, C. et al., 2014. Handwritten Character Recognition by Alternately Trained Relaxation Convolutional Neural Network. In *Proceedings of International Conference on Frontiers in Handwriting Recognition, ICFHR*. pp. 291–296.
Yuan, A. et al., 2012. Offline handwritten English character recognition based on convolutional neural network. *Proceedings - 10th IAPR International Workshop on Document Analysis Systems, DAS 2012*, pp.125–129.
Zhang, Z.H. et al., 2019. Lenet-5 Convolution Neural Network with Mish Activation Function and Fixed Memory Step Gradient Descent Method. In *2019 16th International Computer Conference on Wavelet Active Media Technology and Information Processing, ICCWAMTIP 2019*. pp. 196–199.
Zhong, Z., Jin, L. & Xie, Z., 2015. High performance offline handwritten Chinese character recognition using GoogLeNet and directional feature maps. In *Proceedings of the International Conference on Document Analysis and Recognition, ICDAR*. pp. 846–850.