International Journal on Advanced Science Engineering Information Technology

Vol.8 (2018) No. 4-2
ISSN: 2088-5334

Multi-Classifier Jawi Handwritten Sub-Word Recognition

Anton Heryanto Hasan[#], Khairuddin Omar[#], Muhammad Faidzul Nasrudin[#]

[#] *Faculty Technology and Information Science, Universiti Kebangsaan Malaysia, Bandar Baru Bangi, 43650, Malaysia*
E-mail: anton.heryanto@gmail.com, {ko, mfn}@ukm.edu.my

***Abstract*—The problems and challenges in Jawi handwritten recognition are inherited from Arabic script which consists of cursive natures, large variety of writing styles due to its morphologically rich, ligature, overlapping characters, dialects and the low quality of the manuscripts images. The word segmentation is difficult because the existence of sub words due to the presence of space within words when contain disconnect characters. The performance of previous Jawi handwritten recognition still consider sub-par. There are three main problem of previous approach. First, the recognizer consist of multiple independent components where the improvement of performance in one component not shared across the systems. Secondly, the features extraction using features engineering approach only works on specific subsets of training data and is less capable to handle broader variants of testing data. Finally, the classifier used implicit segmentation where target class is sub-word with limited lexicon. This paper propose use of Deep Learning approach to address the first problem where training is conducted end-to-end from input to class output which enable the improvement of each component to improve overall performance. Secondly, Convolutional Network is use as learning features optimizes the data representation through end-to-end training of the parameters from raw input data to target class. Finally, A multi-classifier implicitly segments the sub-word into sequences of characters are proposed. The classifiers consists of one sub-word length classifier and seven character classifiers. This approach is lexicon-free to address absent of lexicon data. Experiments conducted on a Jawi handwritten standard dataset showed an accuracy of up to 92.20% and suggest that the approach used is superior to state-of-the-art methods of Jawi handwriting recognition.**

***Keywords*— jawi; handwritten recognition; sub-word; end-to-end learning, learning features; convolutional network.**

I. INTRODUCTION

Jawi is subset of Arabic writing used to write Malay language with additional characters to support non-existent phonemes of Malay language. There is a tremendous amount of unexplored Jawi historical handwritten manuscripts which are yet to be studied because it requires Jawi experts which are very few in numbers. Therefore, digitalization and further processing will simplify the information retrieval of the manuscripts.

The problems and challenges in Jawi handwritten recognition are inherited from Arabic script which consist of lots of writing style variants, ligature, overlapping charactes, dialects and the low quality of the manuscripts images [1][2], There are a lot of research on Jawi handwritten recognition but most of the research are conducted on limited lexicon and still show relatively low accuracy[1]. Since Jawi handwriting is unconstraint with large of lexicon. With limited lexicon handwriting recognition is considered unpractical. Therefore Jawi handwritten recognition is still considered open problem.

The cursive nature, presence of ligature and lot of possible writing styles shown by calligraphy introduce more problems in recognizing Jawi handwritten materials. Previous approach to recognize Jawi handwritten materials is by segmenting the artifacts from high level representation on page, line, word into smaller representation of letters[3]. Segmentation into level of word is mostly a solvable problem with state of the art word-spotting technique[8], but in Arabics scripts, the cursive group of characters are not neccessary a complete word but just one part of a large word because the existance of letter which didn't have middle representation. The sub-word is considered as small representation of the Arabic scripts which consists of isolated letter, part of word and the word itself. It poses a different problem where in order to extract meaningful information, the structure predictions for word recognition are required.

Most of the approach can be categorized into explicit segmentation and implicit segmentation[8]. The explicit segmentation requires more components in order to segment the raw image of sub-word into characters, but false positive result will affect the overall performance of systems. Whereas the implicit segmentation provides imaginary segmentation which only focus on overall performance sub-word recognition without taking into consideration the correctness of character segmentation. It is sometimes more

robust on false positive recognition as whole performance is the priority.

Previous research on Jawi handwritten recognizer contains multiple components that handle each step of the process in order to recognize the sub-word [3][4][5][6][7]. This approach depends on the high performance of each component and which most of the time, the improvement of one component does not necessarily improve the overall performance of the recognizer. The training process only conducted on classifier component and only improve classification performances but limit by performance of other components.

The state of the arts of handwritten recognition consist of major components of pre-processing, features extraction, classifier and post-processing components. Previous state-of-the-art Jawi handwritten recognizer only suitable for limited lexicon. The training approaches are mostly on character level because word level training requires more training samples.

The robustness of the features extraction plays major roles to the overall handwriting recognition performances, as the robust features extraction will feed the classifier with informative input to be able correctly classify the handwritten data. But, the major challenge in features extraction is that it only works on specific subsets of training data and is less capable to handle broader variants of testing data. This is because the features extraction is hand-crafted to be able theoretically extract distinct features from the object to be recognized but in reality it requires lots of effort to tune the feature extraction parameters.

Hand-crafted features or feature engineering requires extensive parameter tuning in order to be able to handle wider variety of sample. Simple variants of writing styles, ligatures, and affine transformations of the data will lead to confusing output to classifiers, which lead to worst overall performance of the handwritting recognizers. As the components are independent the improvement at training level on classifier does not propagate to features extraction as the features extraction are hand tuned. Previous Jawi handwritten recognition research made use of feature engineering of geometrical features [3][4][5][6][7][9], but do not really perform well for different varieties of Jawi scripts.

II. Material and Method

The Convolutional Neural Network (CNN) or Convolutional Network (ConvNet) are neural network architectures inspired by mammalian vision cortex[11] which contains features hierarchy of simple, complex and hyper-complex cells. These layers have low-level, medium to high level features respectively. Initial implementation of above architecture is Neocognitron which improved by Convolution Network by adding sub sampling process using max pooling and training the network using back-propagation. ConvNet was specifically designed for handling data with spatial topology (e.g. sound spectrogram, character sequence of text, images, videos or 3D voxel Data)[12] .

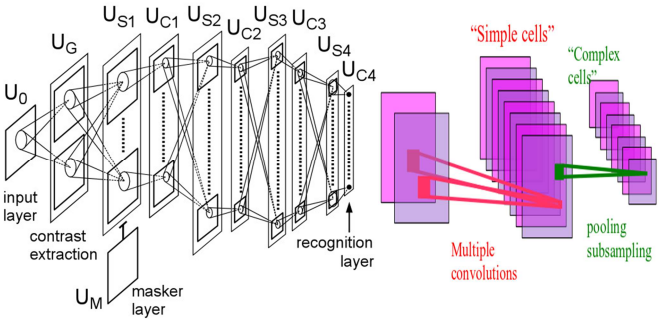

Fig. 1 Neocognition and Convolutional Networks

ConvNet architecture are similar with MLP, which the input layer, hidden layer and output layer, but the hidden layers consist of several components which are convolution layer, non-linearity, pooling layer.

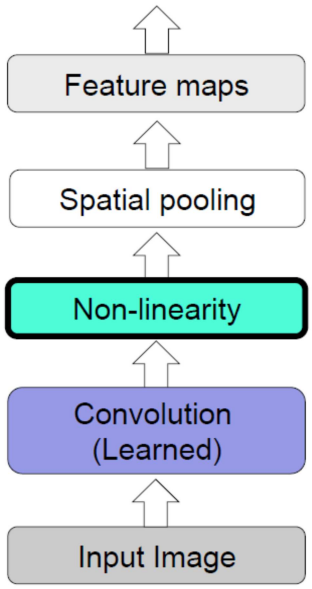

Fig. 2 Convolution Network Layer

A. *Convolution Layer*

The convolutional layer consists of a set of learnable filters or filter banks which produce activations map for each spatial position. It locally connects a hidden layer with a small region of input and share the weight with all neurons to the left and right spatially.

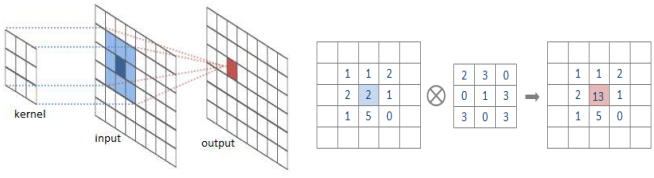

Fig. 3 Convolution operation

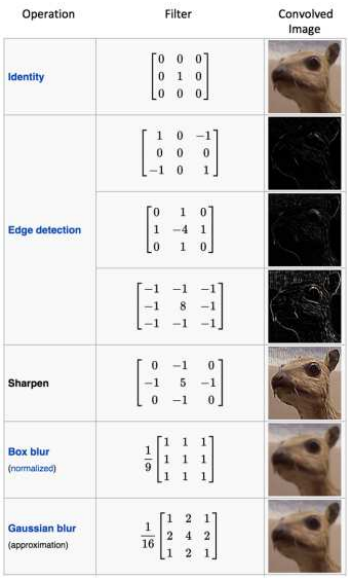

Fig. 4 Convolutional Filter

B. *Non linearity*

Real world data is non-linear, but convolutional operation is linear, therefore to introduce non-linearity to data, the non-linearity operation are used such as Rectified Linear Unit (ReLu), tahn or sigmod, but ReLu is found to perform better in most of situations.

$$f(x) = max(0, x)$$

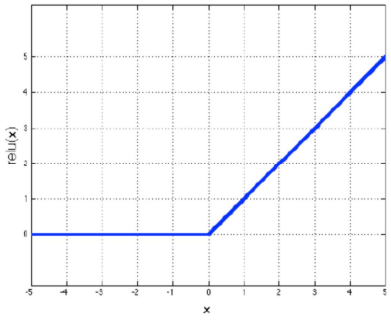

Fig. 5 Rectified Linear Unit

C. *Pooling Layer*

Pooling Layer reduces variant and dimensionality by non-linear down-sampling the spatial dimension (width, height) of the input but retains the most important information, there are several pooling operations such as max, average, sum and etc, but common practices are using max pooling with 2x2 receptive fields and with stride 2. Max pooling take large element from the rectified feature map with the windows on defined receptive field.

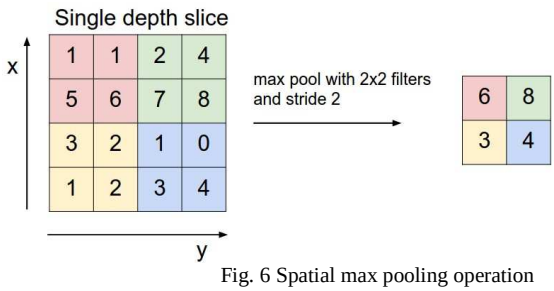

Fig. 6 Spatial max pooling operation

D. *Fully Connected Layer*

The Fully Connected layer is a traditional MLP that uses a softmax activation function in the output layer. The output from the convolutional and pooling layers represent high-level features of the input image. The purpose of the Fully Connected layer is to use these features for classifying the input image into various classes based on the training dataset.

It is considered as a high-level reasoning of ConvNet, apart from classification, adding a fully-connected layer is also a (usually) cheap way of learning non-linear combinations of these features. Most of the features from convolutional and pooling layers may be good for the classification task, but the combination of those features might be even better. The fully connection layer has temporal input layer by flattening the previous layer and connected to hidden layer and output layer as class target and mostly using ReLu for non-linearity.

E. *ConvNet Architecture*

Despite common architecture which consist of two major component which are feature layer and classification layer, there are lots possible combination of component that make of new network architecture of Convolutional Network. This architecture consist of combination of component with different arrangement and parameter setting.

These Architectures consist choice depend on task or application of to handle. Some of these architecture are specific tailored for that specific task, but there are common type architecture which applicable to most of the tasks.

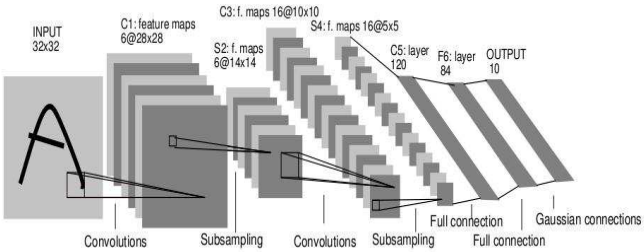

Fig. 7 Early ConvNet Architecture by Lecun named LeNet

Earlier ConvNet Architecure are one that uses for handwritten recognition and consist of two features layer and fully connected networks (FCN) [12]. This architecture quite shallow as training to model to more depth require more computing power and putting more layer reduce the effectiveness of the network. The introduction of several improvement in regularization and training approach enabled the effective uses of more depth approach [13][14].

The implementation of ConvNet using Graphical Processing Unit (GPU) improve the processing performance up to 20 times compare to CPU[15][16]. This improvement further accelerate the introduction of more depth ConvNet Architecture.

AlexNet Winner of ImageNet competition uses similar architecture with LeNet but five features layer and using rectified linear unit (ReLU) for activation function and dropout to improve the training. Which more depth lots parameters and implemented in GPU, this architecture became reference ConvNet implementation [13]. This

architecture improve by GoogleNet by introduction of Inception module which significantly reduce number of parameter[19]. VGGNET introduce more depth for CovNet and shown that its critical component to improve overall performance[20]. Residual Network developed (ResNets) further improve the uses of very depth ConvNet by introducing Residual component with skip connection and not using pooling[21]. Densely Connected Convolutional Network has each layer directly connected to every other layer in a feed-forward fashion [22].

F. Learning Features and Multi-classifier

Using ConvNet as Learning features prove to be successful in handwritten recognition and recently object recognition [10][11][12]. Learning features optimize the data representation of specific tasks. The features parameters are optimized by learning the features end to end with data. Given the advance of Deep learning architecture in which layers of neural could be stacked more than 2 layers and able to represent the features space of object target [13], this improves the back-propagation learning method with dropout and rectified linear (ReLu) unit which improves learning for multiple layers without over-fitting or under fitting the network model [14].

The learning features are able to represent the large variance to training data and are more robust to affine transformations. The ConvNet provides hierarchy representation of the object from low level, middle and high-level representation. The representation of adjusted per training data are very robust according the target tasks. As the nature of ConvNet network is similar to other MLP architecture, it can be trained using back-propagation which can be combined with the classifier into a big neural network. Therefore, the system can be trained as a single system where the improvements of the system is adjusted across the system. The learning features also proven to be quite generative as it can be used for other tasks just by replacing the classifier part of the system.

The propose Jawi Handwritten Sub word recognizer consists of two major components which are Learning Features and Multi-Classifier.

a. Learning Features

Using ConvNet as Learning features has been proven to be successful in handwriting recognition, and recently, object recognition [11] Learning features optimize the data representation of specific tasks. The features extraction parameters are optimized by learning the pattern and representation of hierarchy features from the data. Deep learning architectures such as ConvNet layer can be stacked more than 2 layers and are able to represent the features space of object target [13].

The improvements in regularization using dropout and non-linearity using ReLU contribute to the increase in the representation capability of ConvNet by enabling it to uses more layers for deeper architecture and improved hierarchical representation of data while reducing the possibility of the model becoming over-fitting or under fitting while still trained using back-propagation learning method similar with MLP[14].

In this paper we propose a ConvNet architecture which is suitable as Learning features for Jawi handwritten sub-word recognition. We started from a simple architecture to a more complex architecture but avoided using too complex an architecture as feature space of Jawi image is smaller compared to object recognition problems.

The Input image are resized into 64x64 size. The Learning features part consists of 8 ConvNet Layers, as in Figure 7. Each of ConvNets Layer are consist of convolution layer with kernel size 5x5 and pading 2, then follow by batch normalization layer, then ReLU non-linearity, pooling layer with size 2, stride 1 and padding 1 and finally end with Dropout layer with value 0.2 to connect with subsequent layer. The Learning Features Produce 192 x 7 x 7 = 9408 of input features to Multi-Classifier .

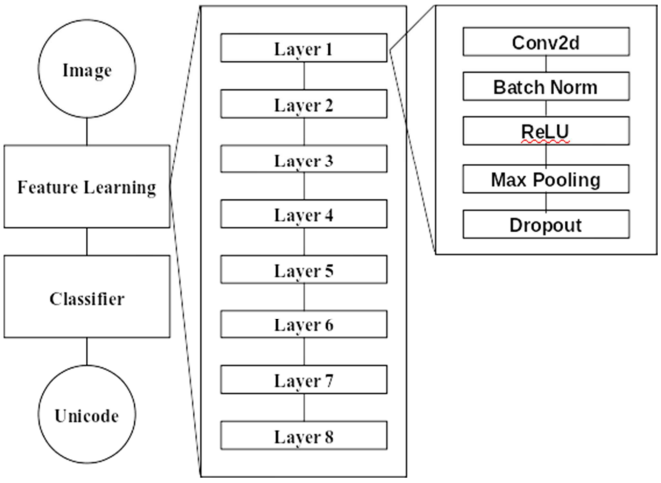

Fig. 8 Multi-layer Features Learning using Convolutional network

In first layer of ConvNet, the convolution layer will process input images with 48 filter banks and produce 48 features maps. This feature map than batch normalize, ReLu and down-sampled using pooling 2x2 as shown by figure 8. and followed by similar component with 64, 128, 160, 192, 192, 192 feature maps size. At the last ConvNet layer, the pooling layer down-sampling the image into 1927x7 features map which flatted into 1x9408 input vector.

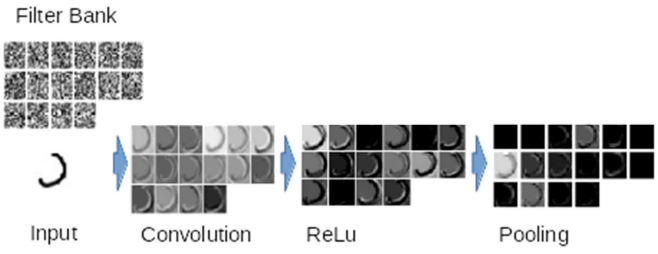

Fig. 9 Show input image in first layer

b. Multi-Classifier

The classification component of Jawi sub-word recognizer, consist of input layer (9408) from learning features layers following ReLu, hidden layer (3072) and final ReLu non-linearity. The hidden layer connect with multi-classifier of Jawi letter and Jawi sub-word length.

There are fix size of classifier with regard to dataset maximum length of sub-word is 10. each classifier target 51 class of jawi letter and symbols. The length classifier to further validate the correct output and improve the semantic capability of ConvNet to implicitly segment the letter in sub-word.

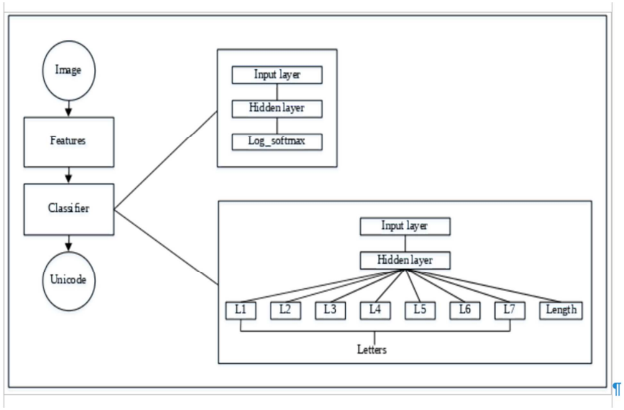

Fig. 10 Multi Classifier sub-word Jawi handwritten recognition

Using multiple classifier. Jawi handwritten recognition will recognize the sub-word by predicting the input length and provides correct sequence of letter using each classifier to determine probability of letter in that position of sequence. Given power of data representation of convolutional network with multiple classifier each letter are implicitly classified into sub-words.

c. Dataset

To evaluate the model, we using Faidzul Jawi dataset[7] which one of standard Jawi dataset to compare previous implmentation.

Fig. 11 Jawi Dataset sample

d. Evaluation

To evaluate the model, we used Faidzul Jawi dataset[x] which is one of the standard Jawi dataset to compare to previous implementations. This dataset is multi-writer Jawi handwritten data based on the new standard of Jawi language. The previous approaches compared to are Trace transform and Circular Coleration with approach and Gabor, ART and SVM approach.

Architecture choice and base line model are also evaluated to analyze the improvement achieved when converting the model from simple MLP to ConvNet and the effects of adding more layers of ConvNets on overall performance.

III Result and Discussion

The following is a comparison of the previous approach with our proposed approach based on Jawi dataset on Sub-word recognition.

TABLE I
Jawi Subword Recognition Accuracy

Methods	**Accuracy (%)**
Trace Transform + Circular Coleration (Limited Lexicon)	60.19
Gabot + ART + SVM (Limited Lexicon)	63.48
MLP (3 Layer) Fully Connected (Limited Lexicon)	70.00
Convolutional Network (3 Layer) + MLP (Limited Lexicon)	81.28
Convolutional Network (8 Layer) + MLP (Limited Lexicon)	90.15
Convolutional Network (3 Layer) + Multi Char (50 Class, 10 Char, Lexicon Free)	71.00
Convolutional Network (8 Layer) + Multi Char (50 Class, 10 Char, Lexicon Free)	92.20

Table 1 shows several types of features and classification methods. The trace transform features by Faidzul[x] performed well on affine transform including rotation and slant invariant, but the output is not really suitable as input of SVM of MLP classifier as position sensitive. The combination of Gabor and ART features using SVM Classifier showed better performance compared to trace transform with circular correlation because SVM are better classifier and Gabor and ART are good features. These combinations are representative of common features engineering and state of the classifier used by features approach.

The neural network approach performs better compared against other approaches. Using MLP with only raw pixels as the input and large parameters, the system performance is better compared to hand-crafted features and kernel method classifiers. The improvements are the effects of the advances of deep learning research which improved the overall neural network approach.

The extensive research on this area in better regularization with dropout, non-linearity with ReLu, and output layer w ith softmax and log softmax. The improvements in propagation learning technique including adaptive learning rate, stochastic gradient decent, small batch, loss function contribute to the overall improvement of neural network architecture. Availability of more powerful processors, larger memory, faster disks with Solid State Drive and advanced usage of Graphic Process Unit (GPU) give the more flexible choice to create better neural network architecture. The availability of deep learning framework open source advance tooling encourage research to further test for suitable architectures.

ConvNet with 3 Layer learning features compared to MLP significantly improved the recognizer performance

from 70 to 81 percent. This performance is for limited lexicon as the output are the sub-word target of dataset which in the real world will be large and not scalable. ConvNet significantly reduces the free parameters compared to MLP by sharing weight and localization features for specific regions of an image. Therefore, it improved the representation of the object being recognized. Further increasing ConvNet from 3 Layer to 8 Layer of Learning features improved performance significantly. The accuracy jumped from 81.28 to 90.15% for an 11% improvement, which highlight that using more layer improves overall accuracy as more hierarchical features are available to provide better data representation.

It is proven to be better at semantically representing the object according to the classification requirements. As the system trained at sub-word level and the improvement of the system back-propagate to features layer and classifier layer, the overall system performance then improved as the training data improved all components of the proposed system. However, it requires large training data in order to improve the learning features generative capabilities and invariant to affine transformation. To further improve the model, training were conducted using mini-batch with adaptive learning rate which enable the model to escape the saddle point after multiple tries.

IV. Conclusions

The improvement of deep learning research improved the architecture availability of neural network which in turn improved the overall neural network approach. The ConvNet is very suitable as learning features for handwritten image. Stacking additional ConvNet Layer will further improve the learning features.

The multi-classifier provides the ability to recognize sub-word in the absence of lexicon, the result shown that it is a very promising approach to handle sub-word recognition of Jawi handwriting. The ConvNet can semantically and implicitly segment the character in Jawi Sub-word and provides correct letter sequence of subword even without knowing the boundary of the letter and with no prior training to segment the letter. The rich deep hierarchical representation of ConvNet provides a way to segment the letter implicitly.

This approach has proven quite simple and effective, but can be further improved using another layer of sequence prediction to improve the overall accuracy and handle more variants of the handwritten data to better generate unknown types of handwriting styles. It is also able to provide better handling of possible out of order sequences and can be further improved with the help of lexicon to correctly recognize sub-word or words.

Acknowledgment

The research is funded by Research Grant FRGS/1/2016/ICT02/UKM/01/1. The research uses hardware from this fund which is DDR4 RAM, Faster Storage m.2 SSD and using Dual GPU Nvidia GTX 1080.

References

[1] M. F. Nasrudin, K. Omar, M. S. Zakaria, C. Y. Liong, 2008, Handwritten Cursive Jawi Character Recognition: A Survey, Proceeding of the 5th International Conference on Computer Graphics, Imaging and Visualization (2008) 247–256

[2] Sitti Rachmawati, S.N.H. Sheikh Abdullah, K. Omar, M.S. Zakaria, & C.Y. Liong, “Review on Image Enhancement Methods of Old Manuscript with the Damaged Background.” International Journal on Electrical Engineering and Informatics. 2(1): 1-14. ISSN 2085-6830. (2010)

[3] K. Omar, “Jawi Handwritten Text Recognition Using Multi-Level Classifier (in Malay),” PhD Thesis, Universiti Putra Malaysia, (2000)

[4] M. Manaf, “Jawi Handwritten Text Recognition Using Recurrent Bama Neural Networks (in Malay),” PhD Thesis, Universiti Kebangsaan Malaysia, (2002)

[5] A. Heryanto, M. F. Nasrudin, K. Omar, "Offline Jawi handwritten recognizer using hybrid artificial neural networks and dynamic programming," in Information Technology, 2008. ITSim 2008. International Symposium, vol.2, no., pp.1-6, (26-28 Aug. 2008)

[6] R. Redika, K. Omar, M. F. Nasrudin, "Handwritten Jawi words recognition using Hidden Markov Models," in Information Technology, 2008. ITSim 2008. International Symposium, vol.2, no., pp.1-5, (26-28 Aug. 2008)

[7] M. F. Nasrudin, M. Petrou, L. Kotoulas, "Jawi Character Recognition Using the Trace Transform," in Computer Graphics, Imaging and Visualization (CGIV), 2010 Seventh International Conference, vol., no., pp.151-156, (7-10 Aug. 2010)

[8] R. Plamondon, S. N. Srihari, 2000, On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22 No. 1 (2000) 63–84.

[9] L. M. Lorigo, V. Govindaraju, 2006, Offline Arabic Handwriting Recognition: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28 No. 5 (2006) 712–724

[10] Azmi, M. S., Nasrudin, M. F., Omar, K., Ahmad, C. W. S. B. C. W., & Ghazali, K. W. M. (2013). Exploiting features from triangle geometry for digit recognition. In 2013 International Conference on Control, Decision and Information Technologies, CoDIT 2013 (pp. 876-880). [6689658] DOI: 10.1109/CoDIT.2013.6689658

[11] D. Hubel and T. Wiesel (1959, 1962, Nobel Prize 1981). Visual cortex consists of a hierarchy of simple, complex, and hyper-complex cells

[12] LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. “Gradient-based learning applied to document recognition.” Proc. IEEE 86, 2278–2324. (1998)

[13] Krizhevsky, A., Sutskever, I. & Hinton, G. “ImageNet classification with deep convolutional neural networks.” In Proc. Advances in Neural Information Processing Systems 25 1090–1098. (2012)

[14] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. J. Machine Learning Res.15, 1929–1958. (2014)

[15] Chellapilla, K., Puri, S., and Simard, P. (2006). High performance convolutional neural networks document processing. In International Workshop on Frontiers in Handwriting Recognition.

[16] Oh, K.-S. and Jung, K. (2004). GPU implementation of neural networks. Pattern Recognition, 37(6):1311–1314.

[17] Hinton, G. E., Osindero, S. & Teh, Y.-W. “A fast learning algorithm for deep belief nets.” Neural Comp. 18, 1527–1554. (2006)

[18] Goodfellow, Ian & Bulatov, Yaroslav & Ibarz, Julian & Arnoud, Sacha & Shet, Vinay. (2013). Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks.

[19] Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition

[20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Er-han, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In CVPR 2015.

[21] Kaiming He and Xiangyu Zhang and Shaoqing Ren and Jian Sun, 2015, arXiv preprint arXiv:1512.03385

[22] Gao Huang and Zhuang Liu and Kilian Q. Weinberger , 2016, Densely Connected Convolutional Networks. ArXiv.1608.06993