JAWI SUBWORD HANDWRITTEN RECOGNITION USING MULTIMODAL SHARED REPRESENTATION: TRACE TRANSFORM NETWORK

Khairuddin Omar, Anton Heryanto, Mohammad Faidzul Nasrudin, Azizi Abdullah, Siti Norul Huda Sheikh Abdullah, and Mohd Zamri Murah.

Centre for Artificial Intelligence Technology,
Faculty of Information Science and Technology,
Universiti Kebangsaan Malaysia
ko@ukm.edu.my

ABSTRACT

This research proposes three main objectives, first the use of Deep Learning approach where training is conducted end-to-end from input to class output, which enable the improvement of each Jawi handwritten recognition component to improve overall performance. Secondly, the use of Trace Transform Network as feature learning to address the features engineering approach by optimizing the data representation through end-to-end training of the parameters from raw input data to target class. This feature learning are more robust to Affine Transformations compared to the state-of-the-arts Convolutional Networks feature learning. Lastly, in order to recognize sub-word, this research proposes a multi-classifier approach, which implicitly segments the sub-word into sequences of characters. The classifiers consists of one sub-word length classifier and seven character classifiers. This approach is lexicon-free to address absent of lexicon data. This research also proposes a hybrid of Trace Transform Network with Convolutional Network feature learning with the advantage of combined robustness of global and local features of the respective networks to further improves the overall performance of the recognizer. Experiments conducted on a Jawi handwritten standard dataset showed an accuracy of up to 93.10% and suggest that the approach used is superior to state-of-the-art methods of Jawi handwriting recognition.

Keywords: Object Signature Features, Convolution Net, Trace Line Net, Trace Angle Net, Trace Line Angle Net, Hybrid Trace Transform Network

1.0 RESEARCH STRUCTURE

The research structure as depicted in Figure 1 composed of three main components. The objectives were formulate as followed:

i. To propose Trace Transform Network feature learning.
ii. To propose lexicon-free Multi-Classifier Jawi Handwritten sub-word recognition
iii. To propose Hybrid Trace Transform Network with Convolutional Network features learning for multi-classifier Jawi handwritten sub-word recognition

The achievement of these objectives as shown in figure 1 reflex the contribution of this research.

Three major focus of this research are identified and shown in research structure. First, Feature Extraction which is major component in handwriting recognition. Second, Sub word recognition is which is unique problem of Jawi and other Arabic scripts descendant. Third is solving the main domain problem in Jawi handwritten recognition.

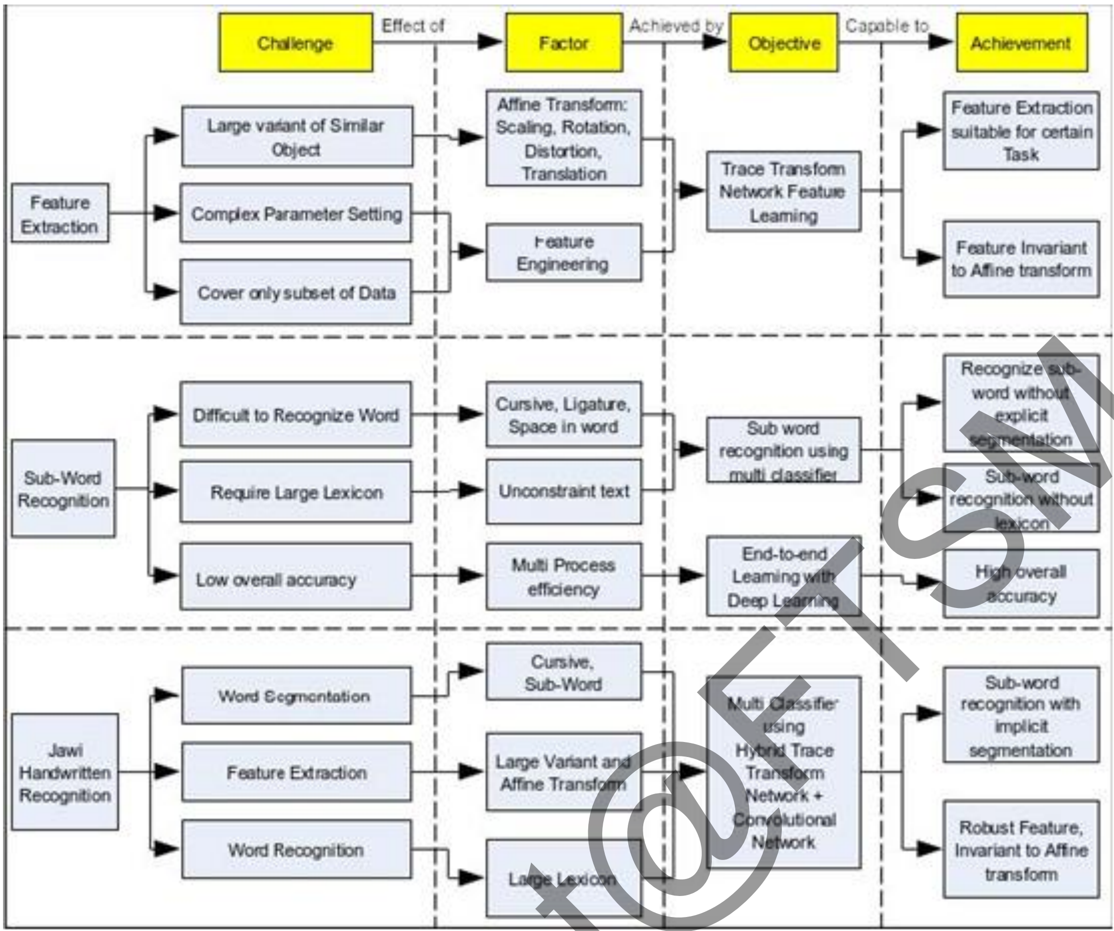

Figure 1: Research Structure

1.1 Features Extraction

The features extraction has the challenges represent the similar object, which has large variant because of the effect affine transformation. This affine transformation including scaling, rotation, distortion and translation. This factor happen because handwriting have variance of medium and writer with different style of writing. Previous research try to handle this problem with features engineering but facing the challenges of complex parameter setting to get robust features. However, even with lots of tuning parameter and evaluation, its only cover subset of data and performance is consider sub par. Therefore, this research propose the Trace Transform Network feature learning. Feature learning extract features from object for optimize for certain task. Using Trace Transform which invariant to affine transformation, feature learning adjust the parameter of Trace Transform to get better parameter according to data thus cover whole subset of data and has generalization capabilities.

1.2 Sub word recognition

Sub word recognition is second focus of this research. This problem cause by nature factor of Jawi scripts which are cursive, has ligature, overlap between character in words and contains space inside the word because of disconnected type of characters. Therefore, instead try isolated and recognise word, the Jawi research focus on sub word recognition and handle word recognition as post-processing. These factors cause word recognition challenges in Jawi handwritten recognition.

Previous Jawi handwritten recognizer using analytical approach try to solve sub word recognition, however facing bigger problem of character recognition which is cause lost information and lower overall accuracy. holistic approach facing problem with large possible lexicon class which causing lower classification result.

Furthermore, large lexicons are requires because un-constraint nature of Jawi handwritten. It further complicate the word recognition.Therefore, this research focus of solving the sub word recognition problem instead of the word recognition and handle word recognition as the next task after sub word recognition.

Previous research propose multiple stages processing and component to overcome this problem. However, each of the component are isolated and evaluate independently thus in-effect and has low overall accuracy of word recognition. This research propose end-to-end learning using Deep Learning approach using multi-classifier. It remove the requirement to do explicit segmentation and improve system parameter from pre-processing until post-processing.

1.3 Jawi Handwritten Recognition

Finally, this research conducted to overcome Jawi handwriting domain problems as it have historical and sentimental value. Jawi handwritten is cursive and has large variance of writing style its require robust word segmentation, features extraction and word recognition. This research propose the robust recognizer by propose robust feature extraction using hybrid feature learning and multi classifier to produce robust Jawi handwritten sub word recognizer. The state-of-the-art Convolutional Network feature learning which is robust local features combine with Trace Transform Network which is global feature learning will handle affine transformation and adversarial problems to better handle variance of Jawi handwritten.

2.0 EXPERIMENT DESIGN

The experimental design is an experimental layout that undertaken by the authors to allows the final output to be analysed and formulated throughout this study. Figure 2 illustrates the experimental design that divided into two main phases; the investigation phase and implementation phase. Each phase has several tasks that are then consolidated into a systematic module. Two initial modules, namely the identification problem and literature review are in the investigation phase.

2.1 Investigation Stages

In this stage, the domain problems are identified and analyzed, and the solutions are proposed based on improvement on previous research and the state-of-the-art approach. Investigations on domain and subjects are conducted in the early phase of the study. In this phase, the background of problems, trends and issues around the domain of issues are explored to get an overview of the research to be made. Factors that lead to problems are also structured. This overview is useful for identifying the subject or topic to explore. Based on the overview, the problem statement and the next research objective are specific. After that, theoretical framework, the importance of the study and the scope of the study are also be formulated. Among the important topics being explored are the general outline of offline Jawi handwritten recognition, strategies in the text recognition system, issues surrounding previous research, feature extraction with feature engineering and feature learning, recognition strategy, deep learning overview and trace transform features.

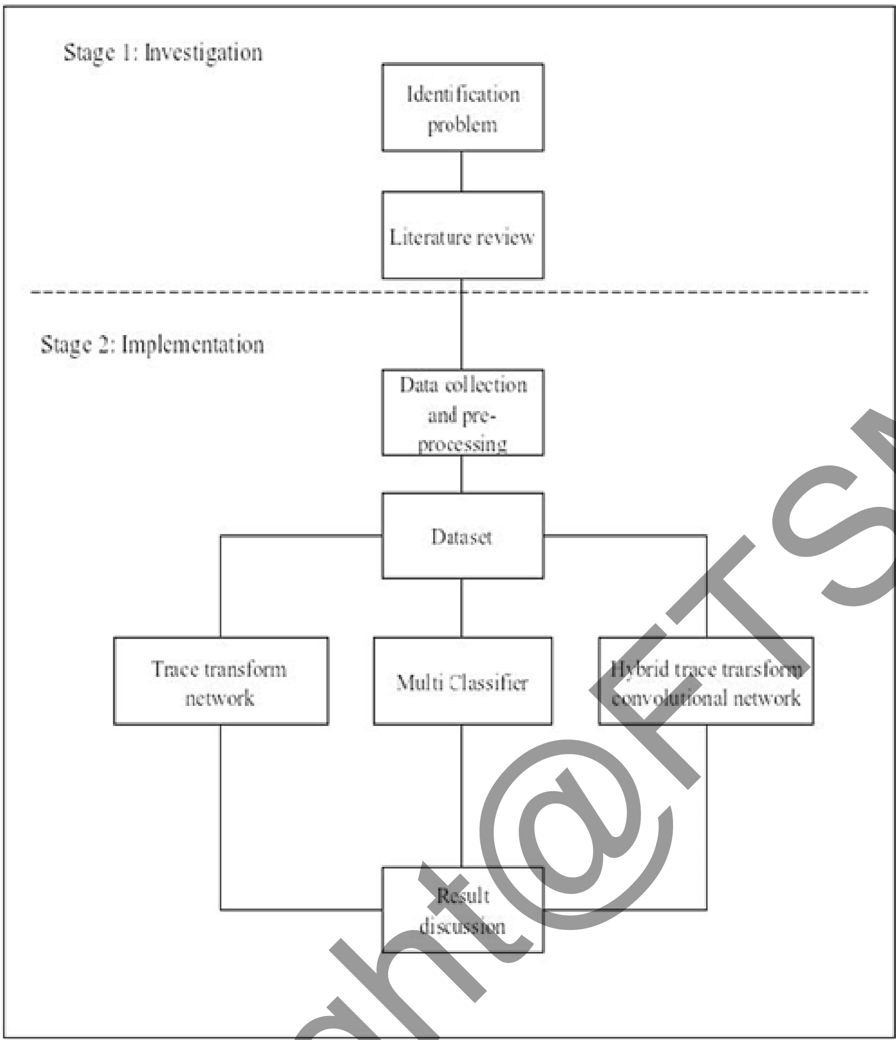

Figure 2: Experimental Design

2.2 Implementation Stage

This phase contains implementation based on experimental design formulated in this research. This stage will discuss about data collection and pre-processing, performance evaluation and Research tools. Following section will discuss detail implementation of Feature learning using Trace Transform Network, sub word recognition using Multi classifier, Hybrid Trace Transform Network with Convolutional Network. Finally, Evaluation of the result, discussion and analysis.

It also means each experiment in each module will be measured and its performance will be analysed. A detailed discussion of each of these modules is included in each section. However, the summary is as follows:

i. Development of Trace Transform Network Feature Learning. In this module, Feature Learning type of Trace Transform Feature are developed based on work on Weight Trace Transform (Srisuk et al. 2006) and inspire by Deep Learning approach in Convolutional Network. Trace Transform Network consist of Trace functional layer which similar with Convolutional layer. This layer will produce feature map in form of sinogram. Diametrical function will act as sub sampling process similar with pooling layer in Convolutional Network. This module were run using algorithm based on (Shin et al. 2008) which uses by Mohammad Faidzul Nasrudin (2010) to generate suitable Trace transform features for Jawi handwritten.

This algorithm generate object signature of Jawi handwritten image using equation in (1).

$$\Pi(f) = \Phi[P[T[f(r,\theta,t)]]], \tag{1}$$

The algorithm for this module given bellow (Anton, 2019):

1. Define 3 parameter of Trace transform:
   - t, distance between subsequent point in each trace line
   - p, distance between trace line in image
   - ϕ, number of trace lines generate in full circle (360 degree).
2. Select the function T, P, and Φ
3. For each ϕ, calculate the trace function, $T = [(r,\theta,t)]$to generate trace transform image.
4. Calculate the diametric function, $P = [[f(r,\theta,t)]]$ to produce object signatures features which generate based on trace line row.

ii. Multi Classifier implementation using only one fully connected network (FCN) classifier for targeted whole sub word in dataset. This multi-classifier used to recognize sub-word (Jawi sub-word recognizer) without explicitly segmented the character from sub-word but instead using implicit approach where character in sub-word recognize by each of sequence of classifier. The classifier will semantically try segmented the sub-word into characters. The classification component of the Jawi sub-word recognizer, consist of input layer, which an output from features learning layers following ReLu, hidden layer and final ReLu non-linearity. The last output layer will consist of fix size of classifier with regard to dataset maximum characters sequence in sub word. Jawi dataset standard maximum length of characters sequence in sub-word is 7. 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 features learning implicitly segment the letter in sub-word.

Using multi-classifier Jawi handwritten recognition will recognize the sub-word by predicting the input length and provides correct sequence of letter using each classifiertodetermineprobabilityofletterinthatpositionofsequence. Given power of data representation of feature learning with multi-classifier each letter are implicitly, classified into sub-words.

iii. Hybrid Trace Transform Network with Convolutional Network explained to discuss the combination both approach which has its own strength. The advance research on Convolutional Network will multiple approaches and architecture, and implementation so it can be use as bases and improved with robust and invariant feature capabilities of Trace Transform Network will improve overall performance of Multi classifier Jawi handwritten sub-word recognition.

3.0 EXPERIMENTAL SETUP

In this research there are three main experiments, Experiments I, II and III as shown in Figure 3. This section describes the details of each experiment in terms of its objectives, interests, inputs, algorithms and outputs. It also describes the role and continuity of each experiment in achieving the objectives of this research.

3.1 Experiment I

i. Objective: Compare the results of the Trace Transform Network, which compose of Trace Layer feature which produce sinogram, and Diametrical layer which produce object signature with previous traditional Trace Transform feature engineering.
ii. Importance: Proving that Trace Transform Network feature learning is the best global feature to use in holistic Jawi handwriting recognition strategy.
iii. Input:Jawi character image.
iv. Algorithm: Experimental Trace functional Layer, Diametrical layer as sub sampling or pooling layer, stacking each layer into Trace Transform Network.
v. Output: A comparative recognition decision between Traditional Trace Transform Features with Trace Transform features learning and Between Trace Trace Transform Network feature learning with Convolutional Network feature learning.

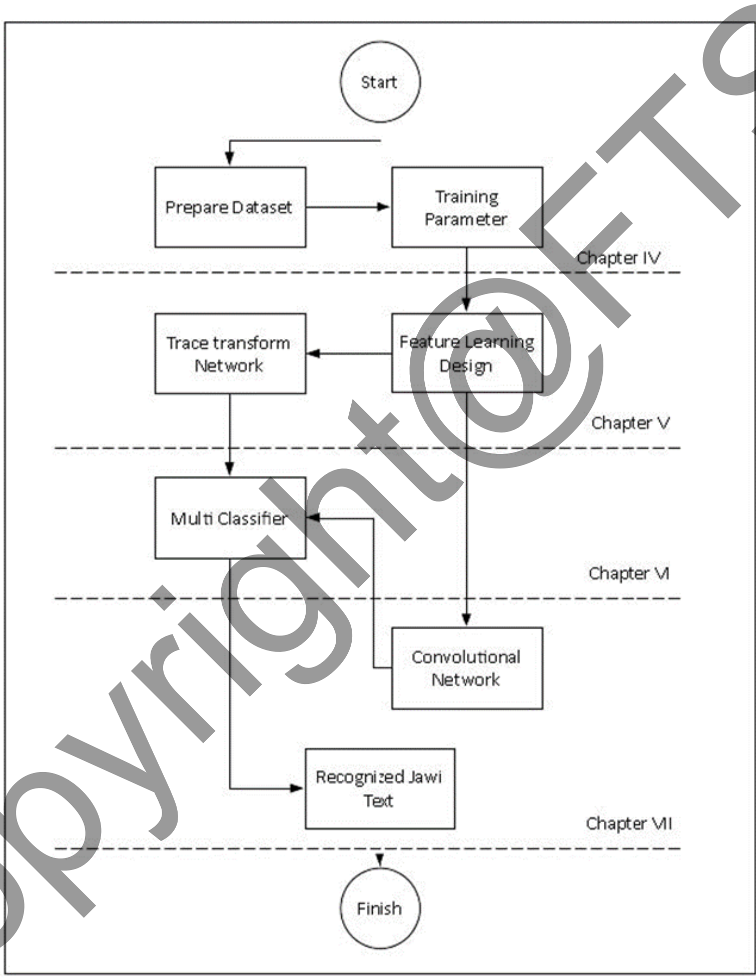

Figure 3: Research Experiment Setup

3.2 Experiment II

i. Objective: To evaluate the effectiveness and robustness of the multi classifier approach compare with previous approach and baseline approach.
ii. Importance: Multi classifier approach using implicit segmentation of character in sub word in order to recognize the sub word. using multi classifier remove the mandatory requirement to uses lexicon in order to recognize sub word, However using lexicon or other post-processing became optional and would further improve recognition accuracy.
iii. Input: Image of Jawi handwritten sub word, Trace Transform feature learning output of Experiment I.

iv. Algorithm: Training parameter and approach, Loss function for multi classifier training and testing.
v. Output: Jawi handwritten sub word, which produce length of sub word and sequence of characters.

3.6.3 Experiment III

i. Objective: Further Improve Feature learning by hybrid the Trace Transform Network with Convolutional Neural Network to produce combination of global and local feature learning.
ii. Importance: produce more robust feature learning by combine local and global feature to improve overall recognition accuracy.
iii. Input: Jawi handwritten subtitle image, Trace Transform Network Feature Learning, Convolutional Network Feature Learning.
iv. Algorithm: Hybrid Convolutional Network (local) and Trace Transform Network (global) feature learning.
v. Output: Robust feature learning approach and best overall Jawi handwritten recognition.

4.0 EVALUATION

Performance metrics for an important recognition model to monitor model progress from a technical point, measure model performance, provide a scientific explanation of model behavior, and identify problems during the model. Performance metrics for optical character recognition (OCR) have been suggested by Kanungo et al. (1999). Kanungo et al. suggested the measure of the performance while testing the famous Arab OCR at that time, OmniPage and Sakhr. According to Kanugo et al. However, OCR performance can be assessed using two approaches i.e. either using a black box or a white box. Black box assessments consider the OCR system as a unit and only the final output of the system will evaluated. On the other hand, white box evaluations assess each output of a system sub module such as distortion correction, page deflation, characterization, classification and so on. In other words, black box evaluations do not take into account the errors of identification caused by sub modules such as sliding.

The performance metrics proposed by Kanungo et al. is the following list assuming *O* is the number of symbols generated by OCR, *M* is the number of symbols that are correctly identified, *D* is the number of symbols lost, *I* is the number of other symbols entered, *S* is the number of symbols in the original text replaced with other symbols, and *T* is the number of symbols in the original text.

- Accuracy, A is defined as. $A = \frac{M}{T}$ The syllabus is also called recall in the field of information retrieval (IR).
- Precision, P is defined as $P = \frac{M}{O}$. Systems with high accuracy are less generating excessive symbols.
- Insert, INS defined as $INS = \frac{I}{T}$.
- Deletion, DEL defined as $DEL = \frac{D}{T}$.
- Replacement, SUB is defined as $SUB = \frac{S}{T}$.

The above performance syllables cannot be fully utilized in this study for several reasons. The first reason is the OCR recognition model is at the character level. On the other hand, the level of recognition of all models in this study is at sub word. In other words, the symbol meant by Kanugo et al. in OCR refers to the characters and symbols in this study are sub word. The second reason is that this study performs manual sub module manually compared to the PAO that automatically executes. So the precision, inserts and deletions of measurement cannot be measured because it relies heavily on the result of deleting pages that you want to recognize. Therefore, only relevant accuracy and substitution are relevant for this study. Given that replacement is in fact the ratio behind the accuracy, the use of precision measures adequately reflects the replacement. In this research, the precision of identification at the rank of i, A_i, defined as:

$$A_i = \frac{M_i}{T} \times 100\% \quad . \tag{1}$$

with M_i is the number of symbols (Jawi subtitle image) in a test data set that is classified accurately as the rank of i and T is the sum of the symbols (Jawi subtitle image) in the test data set. For example, if the measure is for the first rank (i = 1) to the fifth (i = 5), the final value is the sum of the five percentages:

$$P_i = 1,2,3,4,5 = \sum_{i=1}^{i=5} \frac{M_i}{T} \times 100\% \quad . \tag{2}$$

The accuracy of recognition at the first rank is the most important measure of performance for comparison recognition model.

In addition to the differences in the level of recognition, the holistic approach used in this study also makes the output of the recognition model in the rank of a list of similarities between test images and reference images sorted from most to most. This list referred as a rank of similarity. The difference is, in the ranks of several levels of categories used, instead of welding, the category is flat. In addition, in precision, accuracy is more important than retraction, instead of welding, both accuracy and withdrawal are equally important. Examples in rank, the correct ranking on the top 5 is important, but in welding, all welding decisions are equally important. In other words, the accuracy measure at the first rank alone is not sufficient to compare the similarities generated by each identification model.

Based on the above reasons, this study will take on the measure of performance proposed by Geng et al. (2007). They used two long-term measuring techniques in the information retrieval field, Mean Average Accuracy (MAA) (Yates & Neto 1999) and Normalized Deduction Growth (NDG) (Jarvelin & Kekalainen 2002) to determine the performance of a rank. MAA is a measure of the accuracy of the rank assuming that there are two types of items (cross correlation results) that are positive (relevant) and negative (irrelevant).The accuracy of n or A(n), measures the accuracy of the top-of-the-off items defined as:

$$A(n) = \frac{\text{total top positive item} - n}{n} \quad . \tag{3}$$

The average accuracy, AA, defined as:

$$AA = \sum_{n=1}^{N} \frac{K(n) \times pos(n)}{\text{total positive item}} \quad . \tag{4}$$

with n represents the position, N is the number of results, the post(n) is the binary function that determines whether the item at position n is positive (relevant) or negative (irrelevant). In this case, the reference item or image at position n will be positive if it has the same class as the image being tested. Then, multiple-point k-nearest (MPK) is defined as the average point k-nearest (PK) for all items.

NDG is designed to measure the accuracy of the query rank that has some relevant rating ratings. The NDG at position n in the item is defined as:

$$N(n) = Z_n \sum_{j=1}^{n} \frac{2_{R(j)} - 1}{log(1+j)} \quad . \tag{5}$$

where n represents the position, R(j) represents the mark for j rank, and Z_n is the normalization factor to ensure the perfect rank (NDG value) at position n is 1. The final value of NDG is calculated by calculating the average NDG for all queries.

According to Manning et al. (2008), MPK is the sole number that represents the average area under the "precision-recall" curve. This can replace the F-Measure measure, which is the harmonic average for precision and recovery, calculated for all classes. From the point of difference, in NDG the lower the rank of a symbol is recognized the higher the penalty is charged with the proportion of logarithm to its position in the rank. In other words, NDG provides higher values for symbols that are recognized at high ranks such as first and fifth ranks. An important feature is to distinguish two similarity lists that have the same MPK value and can only be distinguished by NDG values.

Both of these metrics are meaningful to compare the strength of an identification model based on the list of similarities in the form of the resulting rank. Therefore, this measure will be used in two early experiments ie Experiments I and II which are more focused on trace sequence. Additionally, these two metrics will be the main topic of discussion in feature selection. Finally, in Experiment III, only accuracy is only used assuming that the user only wants to look at the accuracy of the generated identification rather than the list of similarities. Other than holistic evaluations, assessment the quality of a Jawi handwritten recognition system for analytical approach where sequence of character are considered as performance comparison with previous implementation using similar performance measurement.

For isolated character or sub-word recognition, the mere accuracy (proportion of correctly recognized items) is sufficient. In the applications considered in this research, the transcript is a sequence of words. Counting the number of completely correct sequences is too coarse, because there would be no difference between a sentence with no correct word and another with only one misrecognition.

In a sequence, may contains not only incorrect words, but there might also be inserted or deleted words. Measures such as precision and recall may take these types of errors into account, but not the sequential aspect. The most popular measure of error, used in international evaluations of handwriting or speech recognition, is based on the Levenstein edit distance (Levenshtein, 1966).

This distance counts the number of edit operations required to transform one string into another. The possible edits are:

i. Substitution of one item for another.
ii. Deletion of one item of the sequence.
iii. Insertion of one item in the sequence.

The minimum edit distance between two strings can be retrieved efficiently with a dynamic programming algorithm. The Word Error Rate (WER) is obtained by computing the minimum number of edits from the reference string to the output transcript, normalized by the number of reference words:

$$WER = \frac{n_s ub + n_i ns + n_d el}{n_r ef} \quad (6)$$

where $n_r ef$, $n_s ub$, $n_i ns$, $n_d el$ are respectively the number of words in the reference, and the number of substituted, inserted, and deleted words in the hypothesis. Note that although it is generally expressed in percentage, it may go beyond 100% because of the potential insertions. In this research, the reported WERs are computed with the SClite implementation (Fiscus, 1997).

Similarly, an even finer measure of the quality can be considered to the output sequence in terms of characters, which penalizes less words with a few wrong characters and is less dependent on the distribution of word lengths: the Character Error Rate (CER). It is computed like the WER, with characters instead of words. The whitespace character should be taken into account in this measure, since this symbol is important to separate words. This measure is gaining interest with open-vocabulary recognition systems, which can output potentially any sequence of characters.

$$CER = \frac{n_s ub + n_i ns + n_d el}{n_r ef} \quad (7)$$

5.0 RESULTS

Experimental results in each study development module are cited and analysed. Each module has different approaches and objectives. However, it has the same performance symmetry i.e. the percentage of accuracy of recognition. Percentage accuracy percentage can measure the frequency of images that are recognized or properly categorized. For a holistic strategy, the percentage of accuracy of the identification is a result of similarity measurements between the vector features of the image that you want to capture and the image of all the lexicon entries.

The results of the experimental analysis using this measure of performance can be seen bellow in detail in each experiments.

5.1 Experiment I

To evaluate the performance. The baseline of trace function features using similar function are calculated and using as input for Feed Forward Neural Network (FFNN) with similar architecture. Performance evaluate by training the network to recognize the sub word similar with evaluation of Mohammad Faidzul Nasrudin (2010). The experiment following:

i. Trace Function Features classify using Circular Correlation,
ii. Trace Function Features classify FFNN,
iii. Trace Function Layer using Trace Line Weight,

iv. Trace Function Layer using Theta Weight,
v. Trace Function Layer using Trace Line and Theta Weight, and the result is shown in Table 1 bellow:

Table 1: Result of Trace Function using FFNN (TFNN)

Set	Accuracy	1-5	6-10	11-15	16+
1	73.74	94.23	1.49	0.74	2.05
2	73.18	93.30	2.23	0.56	2.05
3	73.93	93.11	1.86	0.93	2.61
4	74.30	94.23	1.68	0.19	2.23
5	72.81	93.67	2.05	0.37	2.23
6	74.12	93.85	1.68	0.93	2.42
7	74.12	93.67	2.79	0.56	2.42
8	73.74	92.74	1.86	0.56	2.98
9	73.74	93.30	2.23	0.37	2.61
Average	73.74	93.57	1.99	0.58	2.40

From the result shown, simple by using Trace function as input of FFNN improve the performance. By adding weight to trace line further improve the recognition performance. Using theta weight as the best accuracy compare to only trace line weight and combination of trace line and theta weight which shown that rotation information play more significant role compared to the number of trace lines. This result in Table 2, Table 3, and Table 4 also shown that Trace function has very good potential as Feature Learning.

Table 2: Result of Trace Transform Network using Trace Function Trace Weight (TFrw)

Set	Accuracy	1-5	6-10	11-15	16+
1	72.44	92.92	1.49	0.74	2.98
2	73.18	93.30	2.79	0.56	2.23
3	73.74	92.92	2.98	0.93	1.86
4	74.30	93.67	1.49	0.74	2.61
5	73.56	93.30	2.05	0.56	2.61
6	74.67	93.30	2.23	0.74	2.42
7	73.37	92.55	2.98	0.74	2.79
8	73.93	94.23	1.49	0.37	2.42
9	73.37	92.74	2.23	0.93	2.42
Average	73.62	93.21	2.19	0.70	2.48

Table 3: Result of Trace Transform Network using Trace Function Theta Weight (TFtw)

Set	Accuracy	1-5	6-10	11-15	16+
1	75.98	94.23	1.68	1.12	1.49
2	74.67	94.04	2.05	0.93	1.68
3	76.91	94.60	1.30	0.74	2.61
4	75.42	94.04	1.30	0.93	2.42
5	78.03	93.85	2.05	0.93	1.86
6	76.72	94.41	1.49	0.93	2.05
7	76.35	94.41	0.93	1.49	1.86
8	76.72	95.34	0.74	0.74	2.05
9	74.86	95.34	1.86	1.12	1.30
Average	76.18	94.48	1.49	0.99	1.92

Table 4: Result of Trace Transform Network using Trace Function Trace and Theta Weight (TFrtw)

Set	Accuracy	1-5	6-10	11-15	16+
1	72.44	92.74	2.98	1.12	2.61
2	70.58	92.36	2.05	1.49	2.42
3	70.02	93.11	2.42	0.74	2.42
4	71.14	94.04	2.05	0.74	2.42
5	73.37	92.36	2.98	1.49	2.23
6	71.88	92.92	1.68	1.30	2.42
7	73.37	92.74	1.86	1.68	1.86
8	71.14	93.30	2.23	0.74	2.61
9	70.20	92.74	2.23	0.93	2.42
Average	71.57	92.92	2.28	1.14	2.38

Similar with diametrical function, Diametrical layer Calculate the function along the sinogram output of Trace Transform Layer. This layer are has similar concept pooling layer in ConvNet. The Different with original Diametrical function is that, it can choose to calculates diametrical function from group number of trace layer. To evaluate the Trace transform network, the performance of Trace Transform network architecture using trace functions in Table 1 are analyzed. Then compared with previous result as follow:

i. Object Signature Features classify using Circular Cross Correlation,
ii. Object Signature Features classify using FFNN,
iii. Diametrical Layer or Object Signature Layer using Theta Weight.

The following Table 5, Table 6, Table 7, and Table 8 present the accuracy of several features combination define above and presented in % accuracy of the pools.

Table 5: Result of Object Signature 20 Features (OS20)

Set	Accuracy	1-5	6-10	11-15	16+
1	60.71	76.69	6.39	4.32	12.59
2	60.53	77.44	6.02	4.14	12.41
3	60.34	75.56	6.95	3.76	13.72
4	61.09	79.70	5.83	4.32	10.15
5	59.77	80.08	6.77	3.38	9.77
6	59.77	79.89	4.89	2.82	12.41
7	61.84	81.58	5.64	2.26	10.53
8	59.40	77.82	5.64	3.20	13.35
9	58.27	76.32	7.52	4.70	11.47
Average	60.19	78.34	6.18	3.66	11.82

Table 6: Result of Object Signature using Baseline FFNN (OSFN)

Set	Accuracy	1-5	6-10	11-15	16+
1	59.85	65.92	11.55	5.59	12.66
2	60.22	70.20	8.57	5.59	11.36
3	60.60	67.23	9.12	5.21	13.78
4	61.15	69.65	9.87	4.47	11.92
5	60.78	67.04	10.99	4.28	13.41
6	60.97	70.02	9.68	3.54	12.48
7	61.15	67.23	10.80	3.35	13.04
8	60.78	68.90	9.87	5.03	11.55
9	60.04	65.55	9.31	6.15	12.85
Average	60.62	67.97	9.97	4.80	12.56

From the result shown that, using Object signature in FFNN not significantly improve compare with using circular cross-correlation. However, adding theta weight and integrated in as Trace Network improve the performance significantly. Further research using Diametrical Layer as sub sampling are required to evaluated its performance as sub sampling layer.

Traditional Trace Layer features is quite good as shown similar result achieved by Mohammad Faidzul Nasrudin (2010). This shown this features already is good. Further refined will make this further suitable for handwritten recognition. Using Object signature increase performance as more suitable to represent the similarity between object in handwritten recognition

Table 7: Result of Trace Transform Network using Object Signature Theta Weight (OStw)

Set	Accuracy	1-5	6-10	11-15	16+
1	66.74	77.84	7.45	5.21	9.50
2	67.86	78.40	8.19	4.10	9.31
3	68.79	81.19	6.33	3.91	8.57
4	66.37	77.47	8.94	4.10	9.50
5	69.72	78.21	7.26	3.72	10.80
6	69.16	80.45	6.15	3.54	9.87
7	68.23	78.03	5.96	4.84	11.17
8	66.93	77.28	7.08	3.54	12.10
9	69.53	79.89	7.64	3.35	9.12
Average	68.15	78.75	7.22	4.03	9.99

Table 8: Result of Trace Transform Network using Trace Layer and Object Signature (TFDF) Theta Weight

Set	Accuracy	1-5	6-10	11-15	16+
1	69.97	82.96	5.13	7.11	4.80
2	71.09	83.52	5.87	6.00	4.61
3	72.02	86.31	4.01	5.81	3.87
4	69.60	82.59	6.62	6.00	4.80
5	72.95	83.33	4.94	5.62	6.10
6	72.40	85.57	3.82	5.44	5.17
7	71.46	83.15	3.64	6.74	6.48
8	70.16	82.40	4.75	5.44	7.41
9	72.77	85.01	5.31	5.25	4.43
Average	71.38	83.87	4.90	5.93	5.30

In Figure 4, by using The Trace Transform sinogram as direct input of Neural Network shown good result. Integrating the trace layer as special Neural Network layer further improve the performance as its can be train to generalize the feature according to representation learned from data training.

Applying the Diametrical function as sub sampling or pooling shown that its can improve the classification performances. Its also can be used to reduce dimensionality of the features into more compact representation. Adjusting the Diametrical grouping of trace line combined / batched shown that its can be adjusted to different type of tasks. Further Analysis of this effect are required to get the insight.

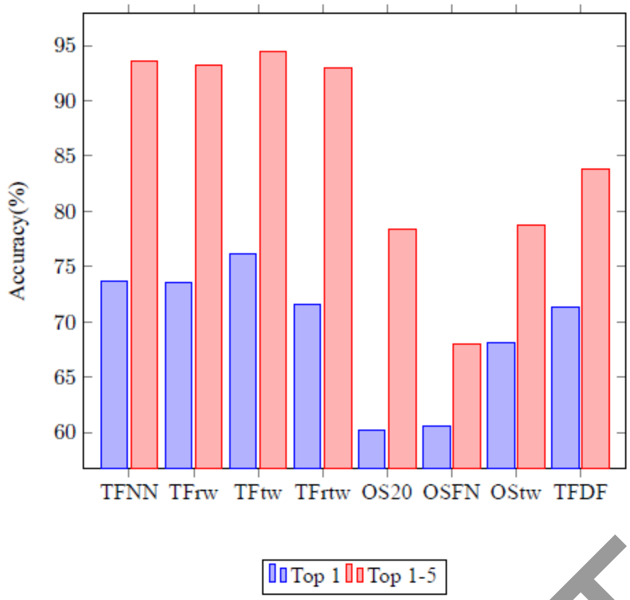

Figure 4: Performance of the proposed Trace Transform Network with previous result.

The result also shows that Trace Transform has potential as Feature Learning, by integrated the Trace Transform operation inside the Deep learning architecture enable to represent data into invariance features which are not sensitive to Affine Transformation such as translation, rotation, distortion, scaling and slanting. This open possibility this features learning used for unsupervised feature learning such as Variable Autoencoder using Trace Transform Networks.

5.2 Experiment II

Experiment Conducted as bellow:

i. Baseline Object Signature Features By Mohammad Faidzul Nasrudin (2010) (OS)
ii. Baseline Object Signature with Multi Classifier (OSMC).
iii. Convolution Network with Multi Classifier 1 Layer (CNMC1).
iv. Convolution Network with Multi Classifier 3 Layer (CNMC3).
v. Convolution Network with Multi Classifier 8 Layer (CNMC8).
vi. Trace Transform Network with Multi Classifier (TTNMC).

The following Table 9, Table 10, Table 11, Table 12, Table 13, and Table 14 present the result of Jawi handwritten recognition of experiment conducted. The result shows the performance of multi-classifiers are better compared to sub-word class target event though has more difficult task to classify sequence of characters. This because the target class of each classifier is smaller and the classifier trained with sub word which consists of single character or more which give the classifier semantic ability to classify the characters.

Based on the result, the multi-classifier is potentially as is produce the sequences of character instead of just sub word class, which give more flexible strategy to extend this framework to improve the result of Jawi sub-word handwritten recognition. Feature learning approach significantly outperform the features engineering approach. The feature learning has more capability to adjust its parameters of features according to data and tasks. Single Layer Trace Transform Network has almost similar performance compared Convolutional Network. This shown that Trace Transform network even though in early state shown potential as feature learning.

Table 9: Result of Object Signature 20 Features

Set	Accuracy	1-5	6-10	11-15	16+
1	60.71	76.69	6.39	4.32	12.59
2	60.53	77.44	6.02	4.14	12.41
3	60.34	75.56	6.95	3.76	13.72
4	61.09	79.70	5.83	4.32	10.15
5	59.77	80.08	6.77	3.38	9.77
6	59.77	79.89	4.89	2.82	12.41
7	61.84	81.58	5.64	2.26	10.53
8	59.40	77.82	5.64	3.20	13.35
9	58.27	76.32	7.52	4.70	11.47
Average	60.19	78.34	6.18	3.66	11.82

Table 10: Result of Object Signature With Multi Classifier

Set	Accuracy	1-5	6-10	11-15	16+
1	68.33	80.32	6.47	9.46	4.80
2	69.45	80.87	7.21	8.34	4.61
3	70.38	83.67	5.35	8.15	3.87
4	67.96	79.94	7.96	8.34	4.80
5	71.31	80.69	6.28	7.97	6.10
6	70.75	82.92	5.17	7.78	5.17
7	69.82	80.50	4.98	9.08	6.48
8	68.52	79.76	6.10	7.78	7.41
9	71.12	82.36	6.66	7.59	4.43
Average	69.74	83.87	4.90	5.93	5.30

Table 11: Result of 1 Layer Convolutional Network with Multi Classifier

Set	Accuracy	1-5	6-10	11-15	16+
1	75.96	87.28	3.02	4.43	5.27
2	77.08	87.84	3.76	3.31	5.09
3	78.01	90.64	1.90	3.12	4.34
4	75.59	86.91	4.51	3.31	5.27
5	78.94	87.66	2.83	2.94	6.58
6	78.39	89.89	1.71	2.75	5.65
7	77.45	87.47	1.53	4.05	6.95
8	76.15	86.73	2.64	2.75	7.88
9	78.76	89.33	3.20	2.56	4.90
Average	77.37	88.19	2.79	3.25	5.77

Table 12: Result of 3 Layer Convolutional Network with Multi Classifier

Set	Accuracy	1-5	6-10	11-15	16+
1	80.43	91.83	3.11	1.29	3.76
2	81.18	92.21	3.02	1.11	3.67
3	81.73	91.83	2.90	1.48	3.78
4	82.29	92.58	2.91	1.29	3.21
5	81.55	92.21	3.67	1.11	3.01
6	82.67	92.21	3.86	1.29	2.64
7	81.36	91.46	4.60	1.29	2.64
8	81.92	93.14	3.11	0.92	2.83
9	81.36	91.65	3.86	1.48	3.01
Average	81.61	92.12	3.45	1.25	3.17

However, the multilayer architecture of convolutional shown significantly improvement over single layer convolution. Increasing the layer from 3 to 8 significantly increase the classification performance, which highlight that using more layer improves overall accuracy as more hierarchical features are available to provide better data representation.

Table 13: Result of 8 Layer Convolutional Network with Multi Classifier

Set	Accuracy	1-5	6-10	11-15	16+
1	83.66	94.16	3.11	1.32	1.41
2	84.41	94.53	3.02	1.13	1.32
3	84.97	94.16	2.90	1.50	1.44
4	85.53	94.90	2.91	1.32	0.87
5	84.78	94.53	3.67	1.13	0.67
6	85.90	94.53	3.86	1.32	0.29
7	84.60	93.78	4.60	1.32	0.29
8	85.15	95.46	3.11	0.95	0.48
9	84.60	93.97	3.86	1.50	0.67
Average	84.84	94.45	3.45	1.28	0.83

Table 14: Result of Trace Transform Network with Multi Classifier

Set	Accuracy	1-5	6-10	11-15	16+
1	77.26	87.62	6.02	3.35	3.01
2	75.95	87.44	6.39	3.16	3.01
3	78.19	88.00	5.65	2.98	3.38
4	76.70	87.44	5.65	3.16	3.75
5	79.30	87.25	6.39	3.16	3.19
6	78.00	87.81	5.83	3.16	3.19
7	77.63	87.81	5.27	3.72	3.19
8	78.00	88.74	5.09	2.98	3.19
9	76.14	88.74	6.21	3.35	1.70
Average	77.46	87.87	5.83	3.23	3.07

Figure 5 shows the comparison of Jawi handwritten recognition based on 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.

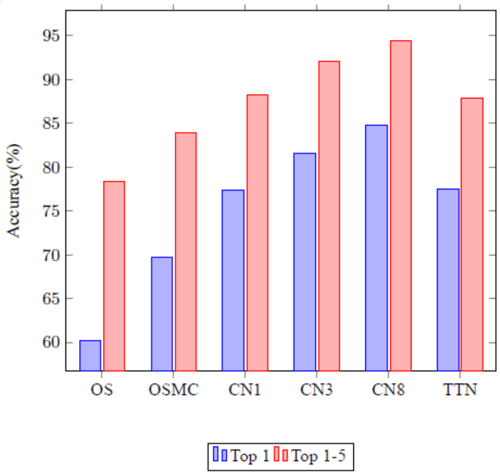

Figure 5: Performance of the proposed handwritten recognition based on the classification requirements.

The propose Trace Transform Network implemented in this research scope only on single layer as further modification on implementation required to make it trainable with backpropagation. Compare with single Convolution Network, the performance are positive and has lots potential. However, the Convolution Network shown that multilayer implementation provide better data representation and further improve the performance of classifier.

5.3 Experiment III

The experiment are conducted in order to evaluate the hybrid strategy to combine the Trace Transform network with Convolutional network. However, the implementation scoped to simple architecture to simplify the analysis. The analysis will explains which approach contributes more to recognition performances. The experiment steps as shown in the Figure 6 conducted to answer following question:

i. Performance when Trace Transform Network placed Parallel with Convolutional Network and connect to Fully Connected Hidden layers.
ii. Performance when Trace Transform Network Layer placed before Convolutional Network Layer.
iii. Performance when Trace Transform Network Layer placed after Convolutional Network Layer.

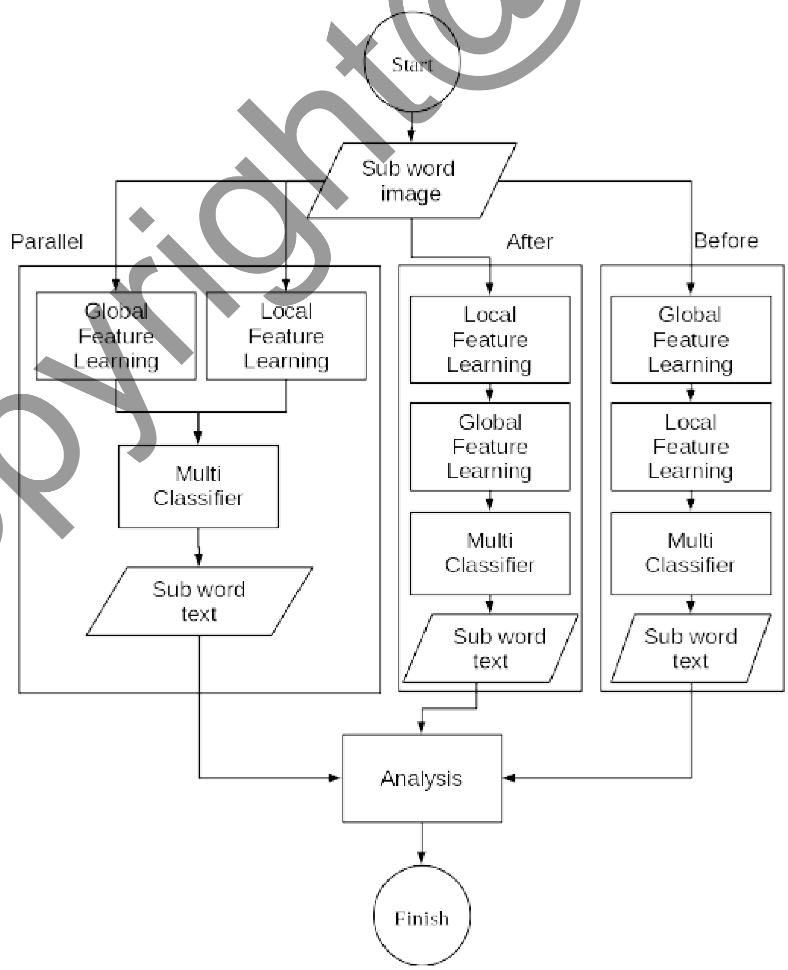

Figure 6: Evaluations approach of Hybrid Feature Learning

The result of each propose architecture are presented similar with previous result tables which consist of the accuracy, the top 5, 5 - 10, 11-15 and the rests. The Table 15 shown parallel hybrid strategy, Table 16 shown the performance result of putting Trace Transform Network before Convolutional Network and finally the Table 17 shown the strategy putting the Trace Transform Network after the Convolutional Network.

Table 15: Parallel arrangement Trace Transform and Convolutional Network

Set	Accuracy	1-5	6-10	11-15	16+
1	87.22	97.47	1.44	0.82	0.27
2	85.92	97.29	1.61	0.93	0.17
3	88.15	97.85	1.17	0.74	0.24
4	86.66	97.29	1.17	0.93	0.61
5	89.27	97.10	1.61	0.93	0.36
6	87.97	97.66	1.35	0.93	0.06
7	87.60	97.66	1.30	0.89	0.16
8	87.97	98.59	0.61	0.64	0.16
9	86.11	98.59	1.43	0.82	−0.83
Average	87.43	97.72	1.30	0.85	0.13

Table 16: Trace Transform placed before Convolutional Network

Set	Accuracy	1-5	6-10	11-15	16+
1	82.07	96.17	0.99	0.74	2.10
2	82.82	96.54	2.29	0.56	0.61
3	83.38	96.17	2.48	0.93	0.42
4	83.94	96.91	0.99	0.74	1.35
5	83.19	96.54	1.55	0.56	1.35
6	84.31	96.54	1.73	0.74	0.98
7	83.01	95.79	2.48	0.74	0.98
8	83.56	97.47	0.99	0.37	1.17
9	83.01	95.98	1.73	0.93	1.35
Average	83.25	96.46	1.69	0.70	1.15

Table 17: Trace Transform placed after Convolutional Network

Set	Accuracy	1-5	6-10	11-15	16+
1	80.90	94.97	2.75	1.12	1.17
2	79.03	94.60	1.82	1.49	2.10
3	78.48	95.34	2.19	0.74	1.72
4	79.59	96.27	1.82	0.74	1.17
5	81.83	94.60	2.75	1.49	1.17
6	80.34	95.16	1.44	1.30	2.10
7	81.83	94.97	1.63	1.68	1.72
8	79.59	95.53	2.00	0.74	1.72
9	78.66	94.97	2.00	0.93	2.10
Average	80.03	95.16	2.04	1.14	1.66

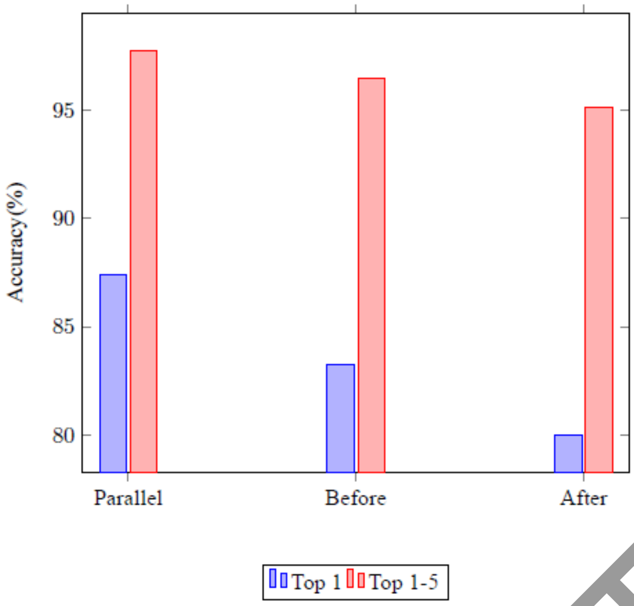

Figure 7: Comparison of the Performance when Trace Transform Network placed before, after, and Parallel.

The above result in Figure 7 shown that the suitable architecture that produced the best result is when Trace Transform network placed parallel or side by side with average accuracy of 87.25% for top accuracy and 97.72% for Top 5 accuracy. This because the both features are contribute to classification performance.

Placing the Trace Transform Network sequential before Convolutional network produce lower result to 83.25% top accuracy and 96.46% on top 5 accuracy compare with parallel approach. The sequential approach by putting Trace Transform network after Convolutional network further lowered the result to 80.03% on top K and 95.16% on Top accuracy. This happen because this strategy produce different set of representation which reduce the robustness of each respective networks. The convolution network provides better result even though the source image is the trace transform images.

Overall result shown that the combination of these features learning is superior compare to either Trace transform network alone and Convolutional Network alone. Therefore, this will put this propose architecture as the new state of the art Jawi handwritten recognizer.

This performance improvement happen because both the feature learning complement each other to produce better representation of data. This combine the global features information and local features information similar with component of visual cortex which have simple cell like convolution dan log-polar cell like trace Transform. The combination of the state-of-art features learning Convolutional Network with Trace Transform Network has potential. It open future research to explorer the used of Trace Transform Network as Invariant Feature Learning for Affine Transform sensitive images.

6.0 CONCLUSIONS

Based on the motivation and issue found the Jawi handwritten research problems which based described by research structure in Figure 1, the Jawi handwritten problem can be solve with end-to-end learning using deep learning approach which enable to classify the sub word into sequences of character by implicitly segment the character inside the sub word using robust representation of features learning.

The improvement in Trace transform feature from feature engineering approach to feature engineering approach, significantly improve the recognition performance of Jawi handwritten recognition compare to previous state-of-the-art. The feature learning approach improve the parameter of feature based end-to-end training from input image sub word to sequence of character classified.

The evaluation on features learning strategy of Trace Transform shown that each weight in trace line or angle has effect on features learning capabilities. Depend on parameter and task to handle this strategies still open to evaluate which one is suitable for given task as it depend on task to handles. In context of Jawi handwritten recognition, using trace line weight or the angle weight is more suitable because combination of weight in trace line and causing more free parameter which require more training data.

Multi-classifier approach try to solve problem in Jawi domain where lexicon available is limited However Jawi historical manuscript is unconstraint with large possible lexicon required. by providing sequence of character as output which is similar with the analytical approach result but still using holistic approach, enable further improvement when the lexicon is available. the architecture choice using deep learning approach enable to extends the network with improvement of injection of prior knowledge or adding the structure predictor to improve the sequence combination.

This research try to produce the new state-of-the-art Jawi handwritten recognizer by hybrid the best local feature learning with extensive research using Convolutional Network with propose Trace Transform Network Global Feature Learning combine with multi classifier. This approach open large of possibility of improvement with more dataset available, improvement on architecture, adding structure predictor and extend with multi modal share representation with Jawi text to improve character sequence combination and produce correct Jawi text not only in sub word level, but further to word, phrase and sentences which enable information retrieval on Jawi historical manuscripts.

The uniqueness of this study lies in the introduction of Trace transform Network features learning which will act as robust global features. This research explore more on the strategy of Trace Transform as feature learning which not explorer by previous research, which are following:

i. The uses of trace function weight not only on trace line but on angle and combination of both which enable more variant of feature representation of the image for given task.
ii. Multi classifier approach shown that possible using feed forward neural network to handle sequence problem in Jawi sub-word recognitions.
iii. The feature learning using Deep learning approach with end-to-end learning which enabled the combination with existing state-of-the-art architecture and techniques.

ACKNOWLEDGEMENT

This paper was supported by Universiti Kebangsaan Malaysia under research grant [FRGS/1/2016/ICT02/UKM/01/1].

REFERENCES

Anton Heryanto, 2019. Jawi Handwritten sub-word recognition based on Traacee Transform Network. Ph.D thesis, Universiti Kebangsaan Malaysia.

Fiscus, J. G. 1997. A post-processing system to yield reduced error word rates: Recognizer output voting error reduction (ROVER). *In IEEE Workshop on Automatic Speech Recognition and Understanding*, pages 347–354.

Geng, Xin, Zhou, Zhi-Hua, & Kate Smith-Miles. Automatic age estimation based on facial aging patterns. TPAMI, 29(12):2234–2240, 2007.

Järvelin, K & Kekäläinen, J. 2002 Cumulated Gaon-Based Evaluation of IR Techniques. ACM Transactions on Information Systems (TOIS) 20, 422-446.

Kanungo, S., Duda S., & Srinivas Y. 1999. "A Structure Model for Evaluating Information Systems Effectiveness". System Research and Bahavioral Science, Vol. 16, pp 495-518.

Levenshtein, V. I. 1966. Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady, 10:707–710

Manning, C.; Raghavan, P.; and Sch ¨utze, H. 2008. Introduction to information retrieval. Cambridge University Press.

Mohammad Faidzul Nasrudin. 2010. Offline Jawi Handwritten Recognition using Trace Transform (in Malay). Ph.D. thesis, Universiti Kebangsaan Malaysia.

Shin, B., Cha, E., Cho, K., Klette, R. & Woo, Y. 2008. Effective feature extraction by trace transform for insect footprint recognition. " in *3rd International Conference on Bio-Inspired Computing: Theories and Applications. IEEE, sep 2008*, pp. 97 – 102.

Srisuk, S., Petrou, M., Kurutach, W. & Kadyrov, A. 2006. A face authentication system using the trace transform 8: 50–61.

Yates, Baeza, Ricardo and Ribeiro-Neto, Berthier. 1999. Modern Information Retrieval. ACM Press, New York. Addison-Wesley.