An Improved Retraining Scheme for Convolutional Neural Network

A. R. Syafeeza[1], M. Khalil-Hani[2], N. M. Saad[1], F. Salehuddin[1], Hamid N.A.[1]
[1,3,4,5]*Universiti Teknikal Malaysia Melaka,*
Hang Tuah Jaya, 76100 Durian Tunggal, Melaka, Malaysia.
[2]*VeCAD Research Laboratory,*
Faculty of Electrical Engineering, Universiti Teknologi Malaysia,
81310 Skudai, Johor, Malaysia
syafeeza@utem.edu.my

***Abstract*— A feed-forward neural network artificial model, or multilayer perceptron (MLP), learns input samples adaptively and solves non-linear problems for data that are noisy and imprecise. Another variant of MLP, known as Convolutional Neural Network (CNN) has additional features such as weight sharing, local receptive field, and subsampling, making CNN superior in handling challenging pattern-recognition tasks. Although CNN has improved the performance of MLP, the complexity of its structure has caused retraining processes to become inefficient whenever new categories or neurons using a winner-takes-all approach are added at the classifier stage. Thus, it is necessary to retrain the complete network set when new categories are added to the network. However, such a retraining incurs additional cost and training time. In this paper, we propose a retraining scheme that could overcome the mentioned problem. The proposed retraining scheme generalizes the feature of extraction layers, hence the retraining process only involves the last two layers instead of the whole network. The design was evaluated on AT&T and JAFFE databases. The results obtained have proved that training an additional category is approximately more than 70 times faster than retraining the whole network architecture.**

***Index Terms*— Convolutional Neural Network, winner-takes-all approach, multilayer perceptron, neural network.**

I. INTRODUCTION

Pattern recognition using computational intelligence approach has been gaining attention for several decades. Computational intelligence approach through neural network learns the data samples adaptively, making neural network very robust toward non-linear data that are noisy and imprecise. The typical neural network is known as a multilayer perceptron (MLP). In pattern recognition approach, MLP is used as a classifier after the feature extraction module. The classifier and feature extraction module are two separate modules in which the designer has to come out with appropriate sets of features in order to classify the samples effectively. In addition, MLP handles one-dimensional (1-D) data, making the pattern recognition system less robust toward changes of input samples.

Another variant of MLP that could overcome the mentioned problem is known as Convolutional Neural Network (CNN), proposed by LeCun et al. [1]. CNN has proved its ability through several applications such as face detection [2, 3], face recognition [4], character recognition [5], gender recognition [6], etc. There are several unique features of CNN. Firstly, it handles two-dimensional (2-D) input samples, meaning CNN accepts raw images and is very robust toward any changes of input patterns such as translation, rotation, and scaling. This deformation is significant due to three architectural concepts applied by CNN, namely weight sharing, local receptive field, and subsampling. Secondly, CNN combines segmentation, feature extraction, and classifier modules into one trainable module. This combination means that the sample features are automatically selected and learned throughout the network. The classifier is normally a set of neurons representing the categories of the patterns. Distance classifier or winner-takes-all rule can be applied to determine the right category of input pattern. Among these two types of classifiers, implementing winner-takes-all rule is simpler than implementing distance classifiers. This is because winner-takes-all rule assigns a winning neuron based on the maximum value obtained among the output neurons, while distance classifier assigns a winner by calculating the shortest distance of a particular set of neurons with the reference image.

Winner-takes-all rule is simple. However, the existence of additional pattern categories can incur additional cost and time. This is because the general CNN in [1] has seven layers and the structure of CNN itself is complex. Therefore, whenever the number of input samples and categories increases, retraining the whole network structure seems to be less practical in the real-world application. As a solution, this paper proposes a retraining scheme that reduces cost and training time effectively. To the best of our knowledge, this is the first proposed work in simplifying the retraining process.

The remainder of this paper is organized as follows: The next section introduces the basic CNN architecture called LeNet-5 and explains the winner-takes-all rule. The third section discusses the methodologies applied. This discussion is followed by the results and analysis and finally the conclusion.

II. THEORY

A. Basic Structure of Convolutional Neural Network (CNN)

LeNet-5 is the basic CNN architecture designed for handwriting recognition application [1]. Figure 1 shows an example of LeNet-5 CNN architecture. It has seven layers, not including the input. Layers C1 through S4 are the feature extraction layers which are formed by alternated convolution and subsampling layers. The final three layers are the general MLP. The output layer is composed of Euclidean RBF units, each one unit representing one category. Each unit accepts 84 inputs from layer F6 and each unit computes the Euclidean distance between the input vector and the weights attached to it. Initially the values of the weights between layer F6 and the output layer are kept fixed at least once and the values follow the stroke of each category (0 to 9 numeral characters) using +1 and -1 values. The training is conducted using Stochastic Diagonal Levenberg Marquardt algorithm [7].

B. Winner-Takes-All Rule

LeCun et al. applies distance classifiers to train the network in order to find the right category. However, there is another approach that is much simpler in terms of computation and could reduce the training time needed. The approach is called winner-takes-all rule.

In winner-takes-all approach, each subject is assigned as an individual category represented by a neuron. During training, a true match is assigned as a label for each input sample. As a result, during the recognition phase, the system will return the identity of the unknown subject.

Winner-takes-all rule is a mechanism that allows a group of neurons to compete with each other as a response to a given set of input patterns. The neurons are similar, but they have different sets of randomly distributed weights attached to them. Each neuron responds differently according to a given set of input patterns, and there is a limit imposed on the strength of each neuron. Only the neuron with the highest activation stays active, while the other neurons are deactivated. The neuron with the highest activation is called the winner-takes-all neuron. This mechanism lies under competitive learning that was introduced by [8]. As the name implies, this concept involves competition among a group of neurons to become active and fire one at a time. This concept is in contrast to Hebbian learning, which has several output neurons that are active at the same time.

The output of a winning neuron, y_k, is set equal to one, while the output of the rest of the neurons is set equal to zero as given in equation 1.

$$y_k = \begin{cases} 1 & if\ f_k > f_j\ for\ all\ j, j \neq k \\ -1 & otherwise \end{cases} \quad (1)$$

Here, f_k refers to the output of the winning neuron k and f_j represents the rest of the neurons.

III. PROPOSED APPROACH

A. Data preparation

There are two types of facial databases used in this work. The first database is AT&T, and it is used to obtain the optimal set of weights. AT&T database was first developed by AT&T Laboratories Cambridge in collaboration with the Speech, Vision and Robotics Group of the Cambridge University Engineering Department in 1992. This database was formerly known as Olivetti Research Laboratory (ORL) database and contains 40 subjects; each subject is photographed with 10 variations in facial expression (open/closed eyes, smiling/not smiling) and accessories (glasses/no glasses), hence a total of 400 images. The files are in PGM format and each of the images is 112×92 pixels with 256 gray levels per pixel. The pre-processing stage for AT&T database is simple, requiring only an image resizing process since the original image is already cropped and it is in the form of grayscale image. There are 40 subjects in the database and the numbers of training and test samples are 320 and 80 respectively, following the 80/20 ratio. The original 112×92 pixel image is resized to 56×46 for subsequent processing. Examples of the image samples prepared for the following experiments are shown in Figure 2.

The second database to evaluate the proposed retraining method is JAFFE, or Japanese Female Facial Expression database. It consists of 10 subjects with 21 samples each. The faces consist of six facial expressions and one neutral face. The images contain frontal image, homogenous background, and various facial expressions. The photos were taken at the Psychology Department in Kyushu University. Figure 3 illustrates some of the samples.

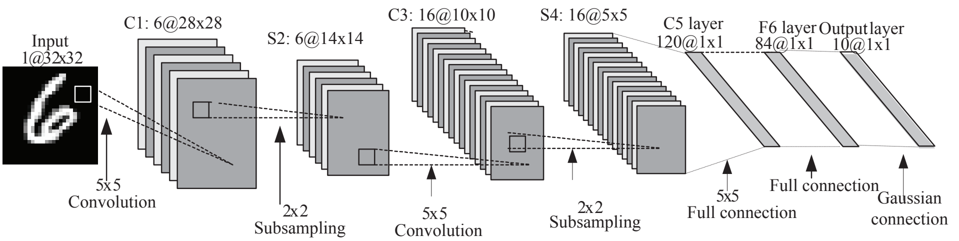

Figure 1: Example of LeNet-5 CNN architecture.

Figure 2: Samples of AT&T database

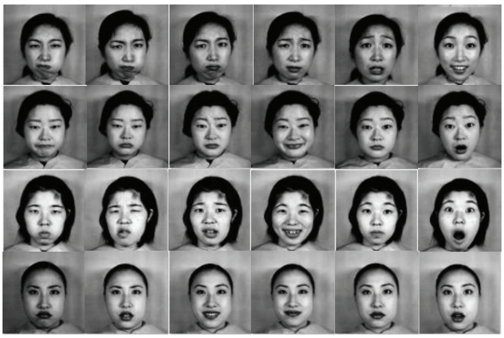

Figure 3: Samples of JAFFE database

B. Methodology

The CNN architecture applied in this paper is a four-layered CNN architecture, in which the convolution and subsampling layers are fused to reduce the number of layers needed as shown in Figure 4. This approach is inspired by Simard et al. in [9]. The detail of designing this four-layered CNN architecture can be referenced in [10]. The type of classifier used in this work is the winner-takes-all approach. Each test sample has been assigned a label which refers to the identity of an individual. In this approach, the highest output magnitude is considered as a winner and the label is compared with the target of the respected sample. If the label and the target is not equal, misclassified sample is counted. Then accuracy is obtained by finding the percentage of the number of correctly classified test samples as given in the following equation:

$$Accuracy = \left(1 - \frac{No.of\ misclassified\ test\ samples}{Total\ no.of\ test\ samples}\right) \times 100\% \quad (2)$$

As mentioned before, winner-takes-all rule is much simpler than distance classifier, but when the number of categories increases, retraining the whole network can incur additional cost and training time. In a normal technique, the network needs to be retrained with the insertion of subjects unseen in the training. Therefore, in this paper, a novel approach involving the last two layers only is proposed by retraining the new subjects. This can be done by first training a group of 40 subjects from the AT&T database. After obtaining the optimal weights, the procedure illustrated in Figure 5 is applied when new subjects are inserted. We evaluated the ability of the proposed retraining scheme using JAFFE database (10 subjects) since this database has similar data complexity to the AT&T database. This means that the feature extraction layers (C1, C2 and C3) are generalized to another type of database. A clear illustration regarding the proposed method can be observed in Figure 6. In the mentioned figure, only one subject is included in the design since there is a total of 41 subjects.

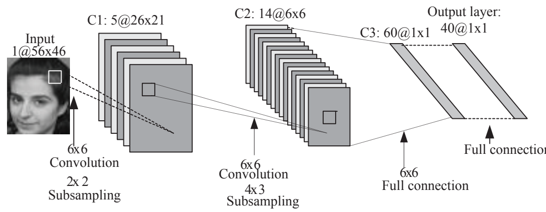

Figure 4: The four-layered CNN architecture with winner-takes-all rule

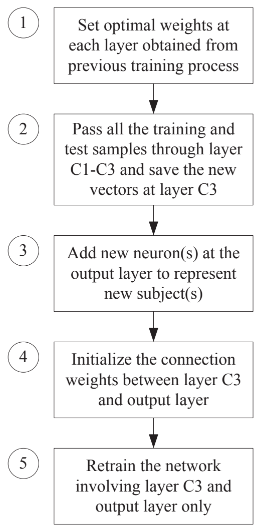

Figure 5: Flow of the proposed method.

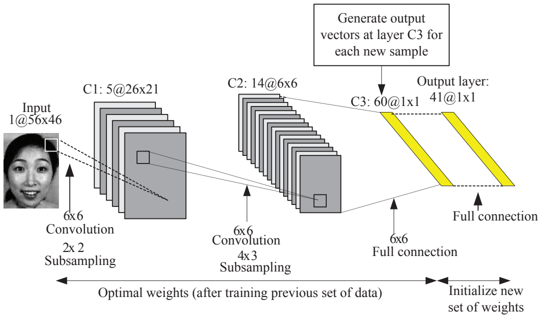

Figure 6: Conceptual view of generalizing the feature extraction layer to the other database in the system.

IV. RESULTS AND ANALYSIS

A retraining scheme is proposed to recognize new categories without retraining the complete set of CNN. This is done by generalizing the feature extraction layers to other databases. In this experiment, the proposed method is evaluated on the AT&T database and new subjects are taken from the JAFFE database.

Each subject in JAFFE database has 21 samples; however, only 10 samples are used for evaluation. As usual, the samples are divided into an 8:2 common training and test ratio. The selected samples from JAFFE are added into the samples from AT&T, making a total of 41 subjects, 328 training samples, and 82 test samples. When 10 subjects are included into the system, there will be a total of 50 subjects which makes a total of 400 training samples, and 100 test samples.

By assuming that the AT&T database has already been trained and optimal weights are obtained, new outputs at layer C3 are generated by passing all the samples into the system. Now the outputs from layer C3 are saved as a new input to the system. A new set of weights is initialized between layer C3 and the output layer before the training is started. The training is conducted until 15 epochs. However, early stopping is set as the stopping criterion which means the training will stop once the misrecognition rate reaches zero.

The result is shown in Table 1. From Table, 3 epochs are needed to train 41 subjects using the proposed retraining method compared to 11 epochs for the normal retraining process. The normal retraining process means that the complete set of CNN is involved to retrain the total number of subjects. On the other hand, 6 epochs are needed to retrain 50 subjects using the proposed retraining method as compared to 11 epochs for the normal retraining process.

Table 1
Result in terms of epochs when accepting new subject(s) from JAFFE database into the system

No. of new subject(s)	Total no. of subjects	No. of epochs for the proposed retraining scheme	No. of epochs for the normal retraining process	Accuracy (%)
1	41	3	11	100
10	50	6	11	100

In terms of the average time taken, the proposed retraining scheme and the normal retraining process for both numbers of new subjects (1 and 10 subjects) have reported prominent results as shown in Table 2. It took about 0.15 and 0.45 seconds to retrain 41 and 50 subjects respectively using the proposed retraining scheme. In contrast, 28.40 and 35.10 seconds were reported for 41and 50 subjects respectively, when using the normal retraining process. These results have proved that the proposed retraining scheme has produced significant improvement in terms of retraining time.

Table 2
Result in terms of average time when accepting new subject(s) from JAFFE database into the system

No. of new subject(s)	Total no. of subjects	Average time to train with the proposed retraining scheme(sec)	Average time to train using normal retraining process (sec)
1	41	0.15	28.40
10	50	0.45	35.10

Figure 7 and Figure 8 depict the result of the training and test errors for 41 subjects and the comparison of mean square error (MSE) between the proposed retraining scheme and the normal retraining process. From the figures, it is obvious that the proposed retraining scheme has produced significantly lower MSE than the MSE resulting from the normal retraining process. Figure 9 and Figure 10 also produced the same result for 50 subjects.

From the results obtained, we can conclude that the proposed retraining scheme produces smaller number of epochs, shorter retraining time and better MSE results than the normal process.

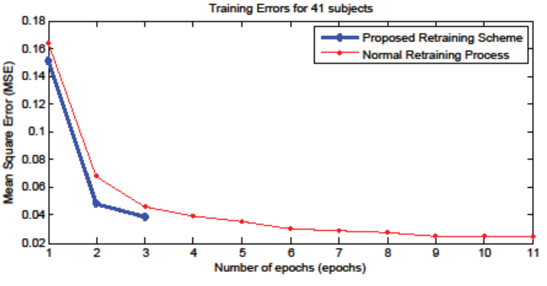

Figure 7: Training Mean Square Error (MSE) for 41 subjects

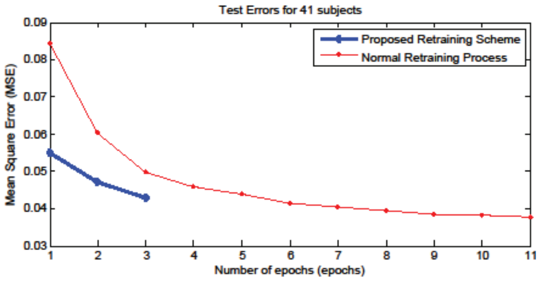

Figure 8: Test Mean Square Error (MSE) for 41 subjects

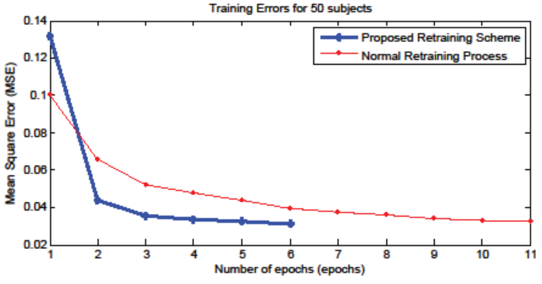

Figure 9: Training Mean Square Error (MSE) for 50 subjects

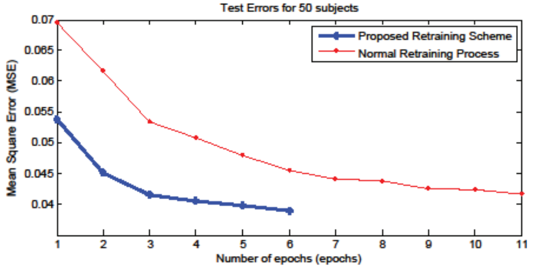

Figure 10: Test Mean Square Error (MSE) for 50 subjects

V. CONCLUSION

In conclusion, using the winner-takes-all rule in conjunction with the proposed retraining scheme is more time effective than retraining the complete set of CNN architecture. The results obtained have proved that the average time taken to train an additional single category has improved to be around more than 70 times faster than retraining the whole network architecture. Therefore, the proposed method has proved to be viable whenever the network grows and increases the number of subjects.

ACKNOWLEDGMENT

This work is supported by Universiti Teknikal Malaysia Melaka (UTeM) through the grant PJP/2014/FKEKK(5B)/S01334.

REFERENCES

[1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," in Proceedings of the IEEE, 1998, pp. 2278-2324.

[2] C. Garcia and M. Delakis, "Convolutional face finder: a neural architecture for fast and robust face detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, pp. 1408-1423, 2004.

[3] F. H. C.Tivive and A.Bouzerdoum, "A face detection system using shunting inhibitory convolutional neural networks," in IEEE International Joint Conference on Neural Networks 2004, pp. 2571-2575

[4] B. Cheung, "Convolutional Neural Networks Applied to Human Face Classification " in 11th International Conference on Machine Learning and Applications (ICMLA), 2012, pp. 580-583

[5] S. A. Radzi and M. Khalil-Hani, "Character recognition of license plate number using convolutional neural network," in Visual Informatics: Sustaining Research and Innovations, ed: Springer, 2011, pp. 45-55.

[6] T. Fok Hing Chi and A. Bouzerdoum, "A Gender Recognition System using Shunting Inhibitory Convolutional Neural Networks," in International Joint Conference on Neural Networks, 2006, pp. 5336-5341.

[7] Y. LeCun, L. Bottou, B. Orr, and K. Muller, "Efficient BackProp," in Neural Networks: Tricks of the trade, Springer, ed, 1998.

[8] D. E. Rumelhart and D. Zipser, "Feature discovery by competitive learning*," Cognitive science, vol. 9, pp. 75-112, 1985.

[9] P. Y.Simard, D. Steinkraus, and J. C.Platt, "Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis," in Seventh International Conference on Document Analysis and Recognition, 2003, pp. 958-963.

[10] A. Syafeeza, M. Khalil-Hani, S. Liew, and R. Bakhteri, "Convolutional Neural Network for Face Recognition with Pose and Illumination Variation," International Journal of Engineering & Technology (0975-4024), vol. 6, 2014.