Learning Face Similarities for Face Verification using Hybrid Convolutional Neural Networks

Fadhlan Hafizhelmi Kamaru Zaman, Juliana Johari, Ahmad Ihsan Mohd Yassin
Faculty of Electrical Engineering, Universiti Teknologi MARA, Malaysia

Article Info

Article history:

Received Mar 19, 2019
Revised Jun 2, 2019
Accepted Jul 28, 2019

Keywords:

Face verification
Face similarities
Unrestricted face
Face recognition
Deep learning

ABSTRACT

Face verification focuses on the task of determining whether two face images belong to the same identity or not. For unrestricted faces in the wild, this is a very challenging task. Besides significant degradation due to images that have large variations in pose, illumination, expression, aging, and occlusions, it also suffers from large-scale ever-expanding data needed to perform one-to-many recognition task. In this paper, we propose a face verification method by learning face similarities using a Convolutional Neural Networks (ConvNet). Instead of extracting features from each face image separately, our ConvNet model jointly extracts relational visual features from two face images in comparison. We train four hybrid ConvNet models to learn how to distinguish similarities between the face pair of four different face portions and join them at top-layer classifier level. We use binary-class classifier at top-layer level to identify the similarity of face pairs which includes a conventional Multi-Layer Perceptron (MLP), Support Vector Machines (SVM), Native Bayes, and another ConvNet. There are 3 face pairing configurations discussed in this paper. Results from experiments using Labeled face in the Wild (LFW) and CelebA datasets indicate that our hybrid ConvNet increases the face verification accuracy by as much as 27% when compared to individual ConvNet approach. We also found that Lateral face pair configuration yields the best LFW test accuracy on a very strict test protocol without any face alignment using MLP as top-layer classifier at 87.89%, which on-par with the state-of-the-arts. We showed that our approach is more flexible in terms of inferencing the learned models on out-of-sample data by testing LFW and CelebA on either model.

Corresponding Author:

Fadhlan Hafizhelmi Kamaru Zaman,
Faculty of Electrical Engineering,
Universiti Teknologi MARA, 40450 Shah Alam,
Selangor, Malaysia.
Email: fadhlan@salam.uitm.edu.my

1. INTRODUCTION

Due to the advancement of deep learning, the quality of image detection and recognition has been increasing for the past five years [1-4]. This also benefits the field of face recognition, where the performance of face recognition has increased by a large margin [5-10]. The key challenges of face recognition in unconstrained environment are variations in poses, illuminations, expressions, ages, makeups, and occlusions. Face recognition task becomes inherently more difficult when faces to be recognized are acquired in the wild.

Traditionally, existing methods generally address the face recognition problem in two subsequent steps namely (1) feature extraction and (2) recognition. In the feature extraction stage, a variation of hand-crafted features has been successfully used [11-14]. Although these works include learning-based feature extraction approaches, each feature are extracted individually and separated from each other, thus some important correlations between the two compared images have been lost at the feature extraction stage.

At the recognition stage, different type of classifiers can be used to either classify each face to its actual identity [15, 16], or just determine its similarity by some distance metrices [17, 18].

Current research in face recognition has produced a few outstanding novel framework or Deep Network architectures. DeepFace network involves more than 120 million parameters using several locally connected layers without weight sharing, rather than the standard convolutional layers [19]. FaceNet uses a deep convolutional network trained to directly optimize the face embedding itself, rather than an intermediate bottleneck layer as in previous deep learning approaches [20]. The benefit of FaceNet is much greater representational efficiency. Liu et al. proposed a two-stage approach that combines a multi-patch deep CNN and deep metric learning, which extracts low dimensional but very discriminative features for face verification and recognition [21]. DeepID3 architectures are rebuilt from stacked convolution and inception layers proposed in VGG net and GoogLeNet to make them suitable to face recognition. Joint face identification-verification supervisory signals are added to both intermediate and final feature extraction layers during training [22]. More recently, Lu et al. proposed a method based on two deep convolutional neural networks (CNN) for face verification and make use of identification signals to supervise one CNN and the combination of semi-verification and identification to train the other one [10].

Nevertheless, there are often misconceptions and misunderstandings on the terms face recognition, face verification, and face identification. Face recognition is a general topic in the field of pattern recognition, which includes both face identification and face verification (sometimes also referred to as authentication). On one hand, face identification is concerned on determining the identity of a person based on the image of the person (client) against all known (labelled) images in databases (galleries). It is basically an answer to the question of "who this person is?". This is also known as one-to-many matching. On the other hand, face verification is focusing on validating a claimed identity based on the image of a client, by comparing the client against a registered image from gallery, whose identity is not necessarily known or labelled. The result of face verification is either accepting or rejecting the claimed identity. This is also known as one-to-one matching. These terminologies are illustrated in Figure 1. In Figure 1, we also show that face verification can be applied as face identification by computing the confidence level, with condition that the label for galleries are known.

One of disadvantage of face verification is good generalization is harder to achieve compared to face identification. However, there are several advantages of face verification employing deep neural networks which includes (1) No retraining required once the network generalize well within scope of training data, (2) Size of network and training data does not change much even when adding new labelled images into galleries for network inference stage, and (3) Can be easily extended to satisfy one-to-many face recognition by use of confidence value. On the other hand, face identification's advantage is it is a relatively simpler approach and good generalization is also relatively easier to achieve. However, there are several disadvantages of face identification deployed in deep neural network environment, including (1) Size of network and training data expands proportional to number of labelled identities and images in galleries, (2) A trained network need to be re-trained when adding new labelled identities and images, and (3) Can be applied to verification case, however, it is highly prone to False Acceptance of Impostor (person not registered/authorized) and False Rejection of Client (registered/authorized person).

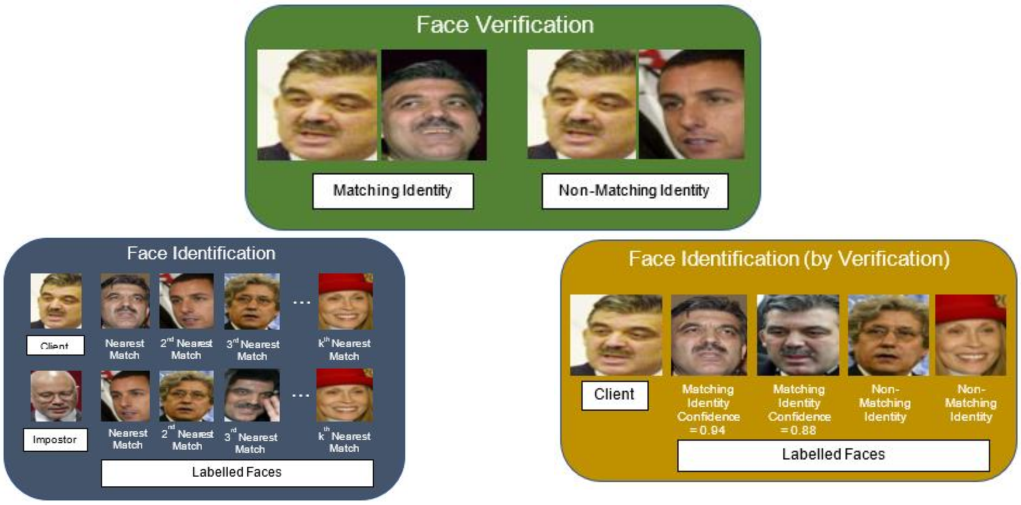


Figure 1. Face Verification and Face Identification terminologies illustrated. Shown together is the implementation of verification as part of Face Identification

Thus, to reduce the hassle of retraining the whole network each time new samples are added, we use face verification approach since it provides more flexibility to the network and database expansion does not require network retraining. It can also further facilitate one-to-many face identification which can be performed simply by taking into consideration the matching identities and their confidence level. The main contribution of this paper is we outline the method to construct 4 different ConvNet which is used to learn similarity of pair of images. These hybrid ConvNet are combined at top-layer level and classified using binary-classifiers. We show the performance of MLP, SVM, Native Bayes and ConvNet classifier in determining whether a pair of images can be considered sharing the same identity or not. We tested 3 different configurations of how the image can be paired together. The paper is organized as follows. Section 2 presents our proposed methods. Section 3 discusses the results obtained in several experiments and Section 4 concludes the paper.

2. RESEARCH METHOD

Previously, hybrid ConvNet approach has been used in [5] where they used 12 ConvNets with 8 combinations of image pairs, arranged in layered configurations. The classifier used was Restricted Boltzmann Machine (RBM). In this work, we propose a much more compact ConvNet with only 8 image pair combinations modeled by 4 separate ConvNets, while assessing the performance of 3 different image pair configurations. The elaboration on each pairing configuration is given in following subsection. Each ConvNet in this work is constructed using the same layer configurations, where the map numbers and dimensions of the input layer and all the convolutional and max-pooling layers are shown in Table 1.

As shown in Table 1, the total layers for each ConvNet is 20 layers, where the first layer is image input layer, taking RGB pair images having i width and j height as inputs (varies according to the size of pair image). There are 7 2D convolutional layers in total (followed by max-pooling layers) which extract the relational features from image pair hierarchically. The map size of each convolutional layers varies from 8 to 512, in ascending order of the hierarchy. Finally, the extracted features pass 3 fully connected layers and are fully connected to 2 neurons in softmax layer which will give the probability of whether the pair image belong to the same person. In this work, we use similar size for Max-Pooling layers which is 2×2 with stride size of 2. For faster computation of activation, instead of sigmoidal activation function, we use a relatively simpler ReLu activation function which is defined as $f(x) = \max(x, 0)$.

Table 1. The detailed parameters of Convolution and Pooling layers.

Layer	Type	Size / Stride	Layer	Type	Size / Stride	Layer	Type	Size / Stride
1	Image Input	$i \times j \times 3$	**8**	2D Convolution	3×3×64	**15**	Max-Pooling	2×2 / 2
2	2D Convolution	3×3×8		Batch Normalization + ReLu		Dropout		
	Batch Normalization + ReLu		**9**	Max-Pooling	2×2 / 2	**16**	Fully Connected	100
3	Max-Pooling	2×2 / 2	**10**	2D Convolution	3×3× 128	ReLu + Dropout		
4	2D Convolution	3×3×16		Batch Normalization + ReLu		**17**	Fully Connected	50
	Batch Normalization + ReLu		**11**	Max-Pooling	2×2 / 2	ReLu + Dropout		
5	Max-Pooling	2×2 / 2	**12**	2D Convolution	3×3× 256	**18**	Fully Connected	2
6	2D Convolution	3×3×32		Batch Normalization + ReLu		**19**	Softmax	2
	Batch Normalization + ReLu		**13**	Max-Pooling	2×2 / 2			
7	Max-Pooling	2×2 / 2	**14**	2D Convolution	3×3× 256	**20**	Top-Layer Classification	2

Each ConvNet is trained using different bootstrap of training data according to their designated pair. When the size of the input regions changes in different ConvNet, the map sizes in the following layers of the ConvNets will change accordingly. To improve the generalization of the ConvNet, data augmentation is employed, where the training data is augmented with random image scaling and XY translations. The output from Softmax layer from all ConvNet are concatenated together to form the final high-level features of the learned face similarity. These features are fed into several types of classifiers to determine whether the learned features belong the same identity or not. The proposed architecture of this hybrid ConvNet is illustrated in Figure 2 while the architecture for each ConvNet is shown in Figure 3.

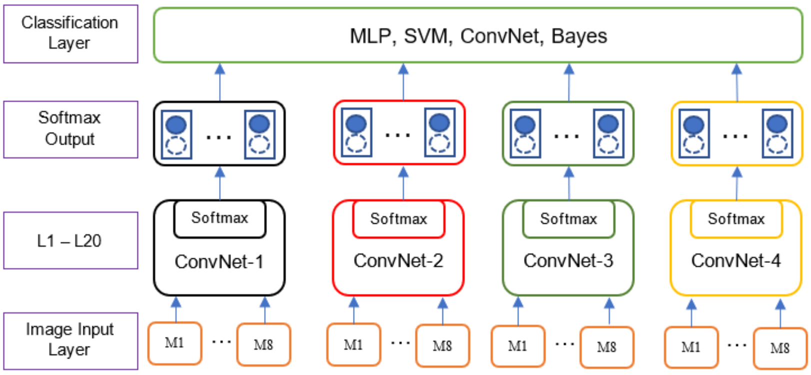


Figure 2. The proposed architecture of the hybrid ConvNet model

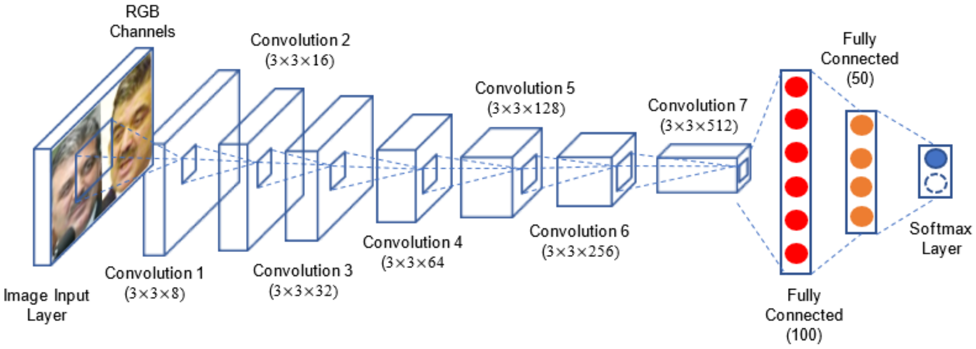


Figure 3. The proposed architecture of the each ConvNet model. For simplicity, only 2DConvolution, Fully Connected and Softmax layers are shown.

As binary classifier, we use 4 binary-classifiers namely Multi-Layer Perceptron (MLP), Support Vector Machine, ConvNet and Naïve Bayes classifiers. These top layer classifiers are used to classify the features from each ConvNet combined. The features of the ConvNet is 2×8 in dimension. Each feature vector belongs to each original face pair is further concatenated in this way, as shown in Figure 4. As a result of this concatenation, each feature vector now has 1×64 dimension for each original face pair.

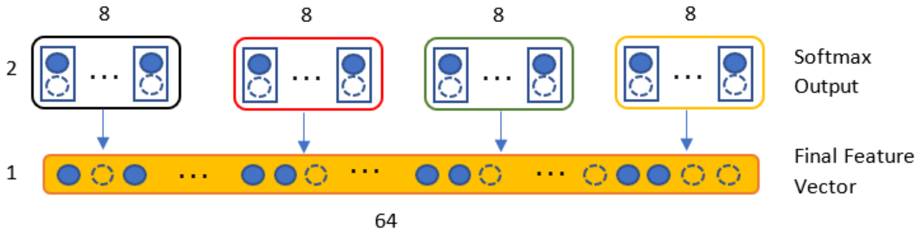


Figure 4. Concatenation of each ConvNet feature to form a final feature vector for each original face pair

MLP classifier used in this work employs scaled conjugate gradient learning algorithm and has 100 hidden neurons with 2 output neurons. SVM classifier and Gaussian is specified as the kernel. Another ConvNet is also used as classifier, where it has a layer of 2D convolution, 3 Max-Pooling layers, and 2 layers of fully connected layers having 50 and 2 neurons respectively. The Bayes classifier on the other hand uses Naïve Bayes with bag-of-tokens model.

We use three different image pairing configurations namely the Lateral configuration, Layered configuration, and Stack configuration. Layered configuration has been used previously where it produces good result [5]. The lateral configuration is constructed by simply combining pair of images horizontally side by side. The layered configuration is made by taking the grayscale intensities of each image, and the average of the two grayscale intensities, and combining them in 3D layers. Let I_{gray} be the grayscale of the first image and J_{gray} be the grayscale of second image, the RGB layered image pair $P_{Layered}$ is denoted as (1):

$$P_{Layered}(R,G,B) = (I_{gray}, J_{gray}, \frac{I_{gray} + J_{gray}}{2}) \quad (1)$$

Meanwhile, the stack configuration is built by combining the pair of images in stack fashion – where first image is vertically placed at the top of second image. These three configurations of image pairing schemes are shown in Figure 5.

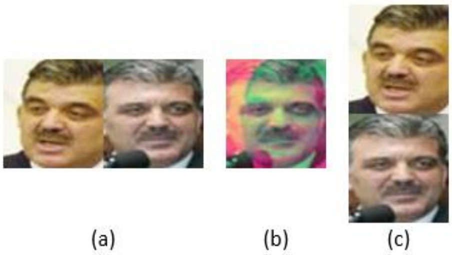


Figure 5. 3 face pair configurations used in this paper: (a) Lateral, (b) Layered, and (3) Stack configurations

For those 4 hybrid ConvNet, each one of them will be used to model the similarity of image pair, constructed by 4 different portions of facial images. The portions are shown in Figure 6. In the meantime, there are 8 different combinations, denoted as M1 until M8, formed from each different portion of face image. The 8 different arrangement of combinations of image pairs used in this work (M1 – M8) are shown in Figure 7. We will also examine the performance of each ConvNet modelling similarity from each face portions to determine the most discriminatively suitable portion of face for face verification.

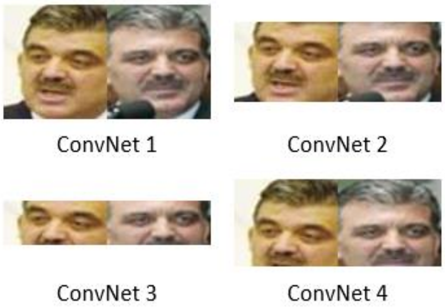


Figure 6. 4 face portions used for each ConvNet

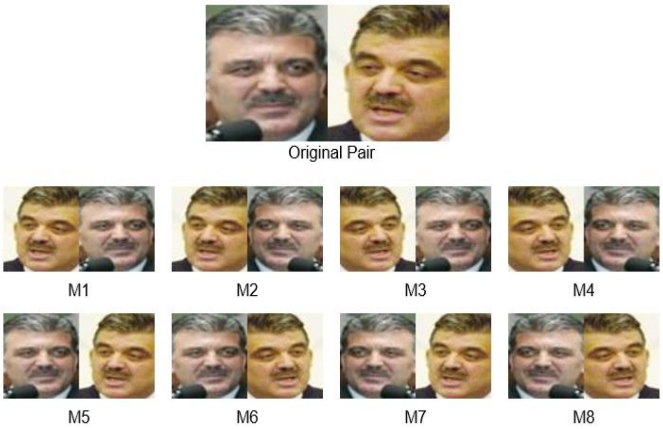


Figure 7. 8 different arrangement of combinations of image pairs (M1 – M8) formed from single original face pair are shown using lateral configurations. Images are flipped and combined to form these combinations

The performance measures used in evaluating this proposed method is by computing the accuracy, True Positive Rate (TPR), False Positive (FPR) and Precision. TPR (also called the sensitivity, the recall, or probability of detection in some fields) measures the proportion of actual positives that are correctly identified as such (e.g., the percentage of image pair having similar identity who are correctly identified as ‘matched’). FPR on the other hand measures the proportion of pair of images having different identity incorrectly classified as having similar identity. TN is the number of pair of images having different identity correctly classified as such, while FN is the number of image pair having different identity who are incorrectly identified as 'non-matched'. Precision is rate of correctly classified matching identity from the whole set classified as having matching identity. The TPR can be computed from (2) while FPR can be calculated from (3). Precision and the overall accuracy can be defined as (4) and (5).

$$TPR = TP / (TP + FN) \tag{2}$$

$$FPR = FP / (FP + TN) \tag{3}$$

$$Precision = TP / (TP + FP) \tag{4}$$

$$Accuracy = (TP + TN) / (TP + TN + FP + FN) \tag{5}$$

3. RESULTS AND ANALYSIS

We evaluate our approach on the well-known LFW dataset containing 7,701 images of 4,281 subjects using the standard image restricted with no outside image protocol [23]. This protocol defines 3,000 positive pairs and 3,000 negative pairs in total and further splits them into 10 disjoint subsets for cross validation. Each subset contains 300 positive and 300 negative pairs. We did not perform any face alignment on the LFW dataset. We also use the CelebFaces Attributes Dataset (CelebA) face dataset which contains 202,599 face images of 10,177 identities (celebrities) collected from the Internet [24]. Following the standard evaluation protocol, images in CelebA and LFW are divided into training and test sets. Furthermore, people in CelebA and LFW are mutually exclusive while the identities in training and test sets are also strictly exclusive. From 10,177 identities in CelebA, 5,901 identities having 2 or more images in the dataset are separated into training and test sets, following the outlined protocol [24]. Each of the 5,901 identities in training and test sets of CelebA are paired with another image with matching identity and non-matching identity. Thus, for each identity, it will have two pairs of images (matching and non-matching). In both LFW and CelebA datasets, 8 image combinations are further formed for each possible face pairs per each ConvNet, thus the total number of pairs in the dataset for each ConvNet is therefore expanded such as described in Table 2. Furthermore, for all four hybrid ConvNets, each will be trained on different face portions formed. In our approach, the training set is further randomly split into training and validation set with 70:30 proportions. We performed two experiments where the first experiment investigates the performance of our proposed method in LFW dataset and compares against some other methods. In second experiment, we perform face verification on CelebA dataset, and compare the performance of face verification on CelebA dataset using face similarity models trained on LFW dataset and vice versa. All experiments are carried out on a computer running on Intel i7-6700 CPU @ 3.40GHz with 16 GB of RAM and GTX 1060 as the main GPU.

Table 2. Number of face pairs used for learning face similarities in each ConvNet

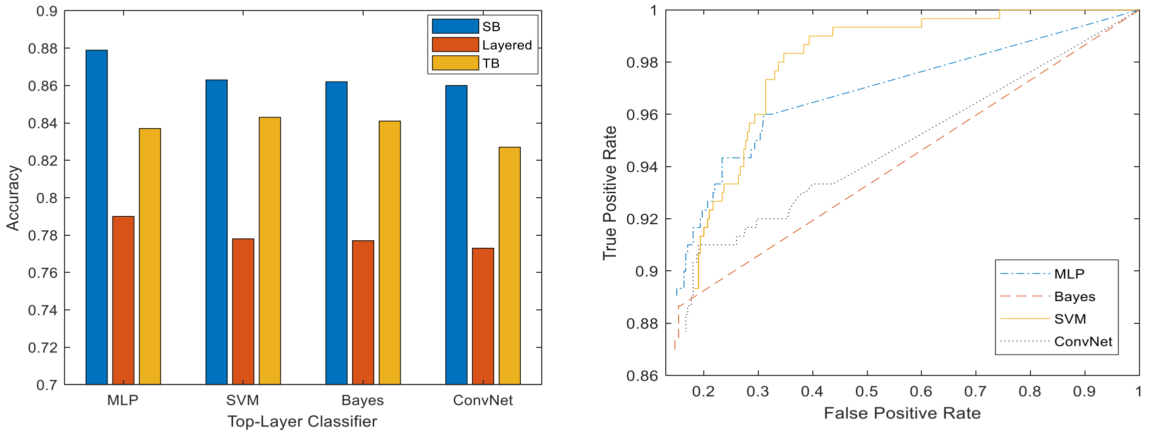
Datasets	# train pairs		# test pairs		Total Pairs
	Positive Pairs	Negative Pairs	Positive Pairs	Negative Pairs	
LFW	86,400	86,400	9600	9600	192,000
CelebA	39,232	39,232	7,976	7,976	94,416

Table 3 shows the performance of face verification on LFW test set using Lateral, Layered and Stack pairing configuration respectively. Besides, performance of each ConvNet models are also shown as separate verification performance. According to the results, Lateral pairing configuration delivers the best accuracy, FPR, TPR and Precision compared to the other 2 pairing configurations. Lateral pairing configuration consistently outperforms Layered and Stack configuration, where it yields 0.879 accuracy using MLP classifier, outperforming others. MLP classifier also delivers best accuracy when compared against other classifiers, where it outperforms SVM best accuracy at 0.863, Bayes' at 0.862 and ConvNet's at 0.860. MLP also yields best FPR, where it produces only 0.137 FPR, the lowest compared against other classifiers. Performance comparison in terms of face verification accuracy between the top-layer classifiers are further shown in detail as bar plot in Figure 8(a).

Table 3. Face verification test performance for LFW Dataset using Lateral pairing configuration

Method		Test Performance											
			Accuracy			TPR			FPR			Precision	
		Lateral	Layered	Stack	Lateral	Layered	Stack	Lateral	Layered	Stack	Lateral	Layered	Stack
ConvNet	ConvNet 1	0.799	0.763	0.790	0.903	0.767	0.827	0.210	0.237	0.213	0.802	0.775	0.798
	ConvNet 2	0.799	0.703	0.788	0.827	0.813	0.820	0.203	0.305	0.215	0.808	0.711	0.796
	ConvNet 3	0.763	0.642	0.574	0.850	0.607	0.590	0.242	0.356	0.427	0.771	0.658	0.589
	ConvNet 4	0.811	0.724	0.778	0.883	0.730	0.777	0.194	0.276	0.222	0.817	0.737	0.789
Top-Layer	MLP	**0.879**	0.790	0.837	0.873	0.790	0.843	**0.137**	0.210	0.170	**0.927**	0.883	0.908
	SVM	0.863	0.778	0.843	0.873	0.787	0.840	0.147	0.230	0.153	0.922	0.871	0.917
	Bayes	0.862	0.777	0.841	0.870	0.789	0.840	0.147	0.230	0.162	0.922	0.881	0.909
	ConvNet	0.860	0.773	0.827	**0.900**	0.803	0.820	0.180	0.257	0.167	0.905	0.858	0.908

Based on Table 3, it is clear that ConvNet 1 consistently delivers the best individual ConvNet performance, while ConvNet 3 is the worst. The results also highlight that the top-classifier approach outperforms individual ConvNet performance. The best improvement in accuracy is achieved when comparing the MLP's accuracy (0.879) against ConvNet 3 (0.763) in Lateral pairing scheme, MLP's accuracy (0.790) against ConvNet 3 (0.642) in Layered scheme, and SVM's accuracy (0.843) against ConvNet 3 (0.574) in Stack Scheme, where the improvements are around 11%, 15% and 27% respectively. Even when comparing between top classifiers' performance against the best individual ConvNet, the improvements are 6%, 3% and 5% respectively which is quite significant. These results point out that the hybrid ConvNet scheme is able to improve individual ConvNet's performance by combining them at the classifier level.



(a) (b)

Figure 8. (a) Face verification accuracy for all 3 different image pair configurations and (b) Receiver Operating Characteristic curve for MLP, SVM, Bayes and ConvNet Top-Layer classifier used in this paper

Further investigation on the results is shown in Figure 8(b) as performance comparison between top-layer classifiers, measured by Receiver Operating Characteristics (ROC) curve. According to Figure 9, as we improve the TPR, SVM emerges as the better classifier compared to other classifiers, as it manages to suppress the rate of FPR from increasing further. Thus, if we look for better TPR to FPR performance, SVM is the best choice. SVM yields 0.99 TPR at 0.3 FPR, surpassing MLP at 0.96 TPR for the same FPR.

Subsequently, the performance of our method is compared against several state-of-the-arts that uses similar restricted image with no outside label protocol as adopted in our work. According to Table 4, our method is on par with state of the arts such as Robust Statistical Frontalization [25], Spartans [26], and Eigen-PEP [27]. When compared against MRF-MLBP [28], our method delivers better results, approximately 9 % better than MRF-MLBP method.

To show that our approach is more flexible in terms of inference of learned model on external data that does not require retraining when adding new out-of-sample images, we perform another experiment. We examine our proposed method on CelebA dataset, where face verification is performed based on the models learned from LFW dataset and vice versa. The results are shown in Table 5. From the results, the accuracy of face verification on CelebA dataset is just slightly affected by the use of LFW model. It decreases from 0.782 to 0.750, however, the TPR is not changed. FPR on the other hand increases slightly too from 0.244 to 0.309. In LFW dataset, the accuracy decreases by 7% when using CelebA model, and similarly, the TPR is not affected with FPR slightly increases from 0.137 to 0.280. The comparison between these models' performance is given as ROC curve in Figure 9. Even though there is slight penalty in accuracy observed, they are still acceptable, and considering that this can remove the hassle of retraining the whole models, this approach is much more flexible and easier to be implemented. Provided that the training data is large enough to model the similarity commonly found in faces, we expect the performance would comparable in the case of out-of-sample inferencing.

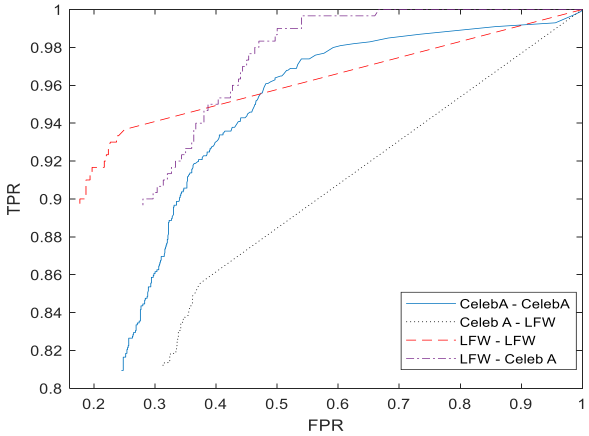


Figure 9. Receiver Operating Characteristic curve for MLP classifier for performance evaluation on inferencing (testing) using different similarity models.

Table 4. Face verification performance comparison against state-of-the arts on LFW dataset with Restricted Image protocol

Method	Accuracy
MRF-MLBP	0.790
Eigen-PEP	0.889
Spartans	0.875
Robust Statistical Frontalization	0.888
Hybrid ConvNet	0.879

Table 5. Face verification performance for LFW and CelebA Dataset using each other's model

(Dataset) – (Model)	Top-Layer Classifier	Test Performance			
		Accuracy	TPR	FPR	Precision
CelebA - CelebA	MLP	0.782	0.809	0.244	0.864
	SVM	0.785	0.812	0.241	0.866
	Bayes	0.788	0.804	0.226	0.874
	ConvNet	0.782	0.832	0.266	0.854
CelebA - LFW	MLP	0.750	0.811	0.309	0.828
	SVM	0.749	0.799	0.300	0.832
	Bayes	0.748	0.787	0.289	0.837
	ConvNet	0.738	0.825	0.348	0.809
LFW - LFW	MLP	0.879	0.873	0.137	0.927
	SVM	0.863	0.873	0.147	0.922
	Bayes	0.862	0.870	0.147	0.922
	ConvNet	0.860	0.900	0.180	0.905
LFW - CelebA	MLP	0.808	0.896	0.280	0.852
	SVM	0.803	0.896	0.290	0.847
	Bayes	0.800	0.896	0.296	0.843
	ConvNet	0.770	0.913	0.373	0.804

4. CONCLUSION

In this paper we propose a framework of our hybrid ConvNet approach to learn face similarity between image pairs for face verification. Face verification is favorable rather than traditional face identification since it can provide us with more flexibility in inferencing trained models on out-of-sample data. We train four individual ConvNet on specific face portions to learn their similarities and combine them into a feature vector at top-layer classifier. The model learns directly and jointly extracts relational visual features from face pairs under the supervision of face identities. 3 different pairing schemes namely Lateral, Layered and Stack configurations are discussed, and 4 different classifiers are used to learn the high-level similarities. We use MLP, SVM, Bayes and ConvNet for this purpose. Based on results obtained, we showed that the hybrid ConvNet approach can improve the performance of individual ConvNet as much as 27%. Even the best performing individual ConvNet can still be improved by 3% when using the hybrid scheme proposed in this work. MLP classifier yield best performance of 0.8789 accuracy on LFW dataset, on par with several state-of-the arts implementing similar test protocol. We found that Lateral pairing scheme delivers the best performance compared to Layered and Stack schemes. We show that the learned model can be applied outside the dataset where the performance penalty is minimal while the implementation will be more flexible. Our proposed approach can be improved further by increasing the number of individual ConvNets and face portions which can enhance the inherent discriminative ability learned by similarity features further. Other classifier such as Joint Bayesian classifier which take the variance of intra and inter-identity into consideration can be used to improve the results further. Jointly training face similarities using verification and identification signals under hybrid architecture can also improve the overall results.

ACKNOWLEDGEMENTS

This research is funded by Institute of Research Management and Innovation (IRMI), Universiti Teknologi MARA (UiTM), Shah Alam, Selangor, Malaysia under the Research Grant Scheme No: 600-IRMI/DANA 5/3/BESTARI (055/2017). The authors also gratefully acknowledge Faculty of Electrical Engineering UiTM Shah Alam for all the support given during this research.

REFERENCES

[1] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 38, no. 1, pp. 142-158, 2016.

[2] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," *ArXiv e-prints*, vol. 1804.02767, 2018.

[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature,* Insight vol. 521, no. 7553, pp. 436-444, 05/28/print 2015.

[4] L. Deng and D. Yu, "Deep Learning: Methods and Applications," *Foundations and Trends in Signal Processing,* vol. 7, pp. 3-4, 2014.

[5] Y. Sun, X. Wang, and X. Tang, "Hybrid Deep Learning for Face Verification," in *2013 IEEE International Conference on Computer Vision*, 2013, pp. 1489-1496.

[6] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1-9.

[7] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Learning Face Representation from Scratch," *ArXiv e-prints*, 2014.

[8] J. C. Chen, V. M. Patel, and R. Chellappa, "Unconstrained Face Verification using Deep CNN Features," *ArXiv e-prints*, vol. 1508.01722, 2015.

[9] Y. Sun, X. Wang, and X. Tang, "Deep Learning Face Representation by Joint Identification-Verification," *ArXiv e-prints*, vol. 1406.4773, 2014.

[10] X. Lu, Y. Yang, W. Zhang, Q. Wang, and Y. Wang, "Face Verification with Multi-Task and Multi-Scale Feature Fusion," *Entropy*, vol. 19, no. 5, p. 228, 2017.

[11] H. V. Nguyen and L. Bai, "Cosine similarity metric learning for face verification," presented at the Proceedings of the 10th Asian conference on Computer vision - Volume Part II, Queenstown, New Zealand, 2011.

[12] F. Kamaruzaman and A. A. Shafie, "Recognizing faces with normalized local Gabor features and Spiking Neuron Patterns," *Pattern Recognition,* vol. 53, pp. 102-115, 2016.

[13] W. J. Li, J. Wang, Z. H. Huang, T. Zhang, and D. K. Du, "LBP-like feature based on Gabor wavelets for face recognition," *International Journal of Wavelets Multiresolution and Information Processing,* vol. 15, no. 5, Sep 2017, Art. no. 1750049.

[14] W. Li, Y. Wang, Z. Xu, Y. Jiang, Z. Lu, and Q. Liao, "Weighted contourlet binary patterns and image-based fisher linear discriminant for face recognition," *Neurocomputing,* vol. 267, no. Supplement C, pp. 436-446, 2017/12/06/ 2017.

[15] F. K. Zaman, A. A. Shafie, and Y. M. Mustafah, "Robust face recognition against expressions and partial occlusions," *International Journal of Automation and Computing,* journal article vol. 13, no. 4, pp. 319-337, 2016.

[16] H. Li, F. Shen, C. Shen, Y. Yang, and Y. Gao, "Face recognition using linear representation ensembles," *Pattern Recognition,* vol. 59, pp. 72-87, 11// 2016.

[17] M. Guillaumin, J. Verbeek, and C. Schmid, "Is that you? Metric learning approaches for face identification," in *2009 IEEE 12th International Conference on Computer Vision*, 2009, pp. 498-505.

[18] C. Huang, S. Zhu, and K. Yu, "Large Scale Strongly Supervised Ensemble Metric Learning, with Applications to Face Verification and Retrieval," *ArXiv e-prints,* vol. 1212.6094, 2012.

[19] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1701-1708.

[20] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A Unified Embedding for Face Recognition and Clustering," *ArXiv e-prints,* vol. 1503.03832, 2015.

[21] J. Liu, Y. Deng, T. Bai, Z. Wei, and C. Huang, "Targeting Ultimate Accuracy: Face Recognition via Deep Embedding," *ArXiv e-prints,* vol. 1506.07310, 2015.

[22] Y. Sun, D. Liang, X. Wang, and X. Tang, "DeepID3: Face Recognition with Very Deep Neural Networks," *ArXiv e-prints,* vol. 1502.00873, 2015.

[23] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller., "Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments," in "Technical Report 07-49," University of Massachusetts, Amherst2007.

[24] Y. Sun, X. Wang, and X. Tang, "Deep Learning Face Representation from Predicting 10,000 Classes," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1891-1898.

[25] C. Sagonas, Y. Panagakis, S. Zafeiriou, and M. Pantic, "Robust Statistical Frontalization of Human and Animal Faces," *International Journal of Computer Vision,* vol. 122, no. 2, pp. 270-291, 2017/04/01 2017.

[26] F. Juefei-Xu, K. Luu, and M. Savvides, "<italic>Spartans</italic>: Single-Sample Periocular-Based Alignment-Robust Recognition Technique Applied to Non-Frontal Scenarios," *IEEE Transactions on Image Processing,* vol. 24, no. 12, pp. 4780-4795, 2015.

[27] H. Li, G. Hua, X. Shen, Z. Lin, and J. Brandt, "Eigen-PEP for Video Face Recognition," in *sian Conference on Computer Vision (ACCV)*, 2014.

[28] S. R. Arashloo and J. Kittler, "Efficient processing of MRFs for unconstrained-pose face recognition," in *2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2013, pp. 1-8.

BIOGRAPHIES OF AUTHORS

Fadhlan Hafizhelmi Kamaru Zaman received the B.Sc (Hons.) and P.hD. degrees from International Islamic University Malaysia in 2008 and 2015, respectively. He is currently a Senior Lecturer at Department of Computer Engineering, University of Technology MARA, Malaysia. His research interests are in surveillance system, pattern recognition, signal and image processing, artificial intelligence and computer vision. Fadhlan is also a member of IEEE, Malaysian Board of Technologist (MBOT), and a Chartered Engineer from the Institution of Engineering and Technology, UK.

Juliana Binti Johari is an Associate Professor in Instrumentation and Control System Engineering at the Faculty of Electrical Engineering (FKE), Universiti Teknologi Mara (UiTM). She received her Ph.D. in Microengineering and Nanoelectronics from Universiti Kebangsaan Malaysia, MSc in Biomedical Engineering from University of Surrey, United Kingdom and B.Eng. (Hons.) in Electrical and Electronics Engineering from University of Strathclyde, United Kingdom. Her research interests are in Microelectromechanical Systems (MEMS), Microfluidics, Biomedical Engineering, Artificial Intelligence and Engineering Education. Juliana is also a Senior Member of Institute of Electrical and Electronics Engineers (IEEE) and a Chartered Engineer from the Institution of Engineering and Technology (IET), United Kingdom.

Ahmad Ihsan Mohd Yassin received his B.Eng. degree in Electrical Engineering from the Universiti Tun Hussein Onn Malaysia. He obtained his MEng and PhD from the Universiti Teknologi MARA, Malaysia. He is currently a Senior Lecturer in the Faculty of Electrical Engineering, Universiti Teknologi MARA. His research interest includes neural network, deep learning, system identification, optimization and blockchain technology. Dr Ihsan is also a senior member of IEEE, a Professional Engineer of the Board of Engineers Malaysia and a Chartered Engineer from the Institution of Engineering and Technology.