Transfer Learning for Makeup Face Verification

Kim-Uyen Nguyen-Thi
Department of Mechanical and Computer Aided Engineering
Feng Chia University
Taichung, Taiwan
ntkuyen0410@gmail.com

Jyun-We Huang
Department of Mechanical Engineering
National Cheng Kung University
Tainan, Taiwan
t9988776@gmail.com

Hsien-Kuang Liu
Department of Mechanical and Computer Aided Engineering
Feng Chia University
Taichung, Taiwan
hkliu@fcu.edu.tw

Shih-Hung Yang*
Department of Mechanical Engineering
National Cheng Kung University
Tainan, Taiwan
vssyang@gs.ncku.edu.tw

***Abstract*—Wearing various styles of makeup makes face verification task difficult. The existing face verification model can extract discriminate features of European faces while it weakly extracts discriminate features of Asian faces. This makes verification of Asian face with makeup more difficult. This study proposes a transfer learning scheme for makeup face verification. A pre-trained Dlib model is fine-tuned with a proposed dataset consisting half of Asian and half of European identities. The pre-trained model was fine-tuned by the proposed dataset in order to correctly verify Asian faces as good as European faces. The experimental results show that the fine-tuned model achieved highest accuracy on existing makeup datasets compared to other works. The results also show that the transfer learning scheme makes the model robust to makeup in face verification task.**

Keywords*— *face verification, makeup face verification, fine-tune, transfer learning

I. INTRODUCTION

Research on human face is a traditional research from the beginning of computer science, especially in computer vision field. Face recognition is an important research topic and involves two main tasks: face identification (to identify the identity image in dataset) and face verification (to verify whether the face belongs to the identity). A recently famous proposed model FaceNet [1] solved face recognition problem by using deep convolutional neural network (CNN) with efficient triplet loss function.

One of the face identification problem is the effect of makeup which might change representation of identity and result in low identification accuracy. The makeup layout makes the outlook of a nose thinner, higher, and smaller by contouring nose makeup technique. It also makes the outlook of the eyes bigger and more attracted. These makeup techniques change the outlook of face and make face recognition difficult.

Few style transfer and learning correlation-based methods have been proposed to deal with makeup face verification. However, they require paired makeup and non-makeup images. Furthermore, some makeup styles may be unseen in training data. Two famous recognition models, Facenet and Dlib [2], have been proposed to extract features robust to makeup. However, these two models pre-trained from European faces may not perform well for Asian faces. They might not discriminant positive and negative pairs well. This study proposes a transfer learning scheme to fine-tune the model so that the model could achieve high accuracy in both Asian and European faces and solve the problem of imbalance training data between Asian and European faces.

II. RELATED WORKS

Three approaches have been proposed to solve makeup face verification, which are style transfer, learning correlation, and transfer learning. Style transfer based methods, such as PairedCycleGAN [3], BLAN-1 [4], BLAN-2 [5], are developed based on concepts of CycleGAN [6] which learn a mapping from makeup to non-makeup, and then use the de-makeup face for verification. Learning correlation-based approach learns the correlation between makeup and non-makeup by a data augmentation method [7]. One study found that makeup has some standard styles on local structure, then learns a pairwise dictionaries for face regions (skin, lips, and eyes) before and after makeup [8]. Another study conducts a correlation mapping between local patches rather than facial features [9]. Transfer learning based methods design a pre-trained model trained from the face datasets collected from internet, then fine-tune the network on a small makeup dataset [10].

A. Style Transfer Based Method

Transferring makeup image into non-makeup image may reserve facial features of the identity which makes the identification robust to makeup. CycleGAN is a well-known model for transferring an image in one domain to another domain without requiring a pair of images. Several previous works utilized the advantages of CycleGAN to remove makeup before extracting face features. Most recently, a PairedCycleGAN [3] is proposed to learn to transfer source image from non-makeup domain to makeup domain referring to a reference image. It also learns to transfer makeup image into non-makeup image. Different from CycleGAN, PairedCycleGAN generates sharp image due to the use of a discriminator. PairedCycleGAN could transfer non-makeup image into various styles of makeup. However, it does not work well on extreme makeup styles which is unseen during training.

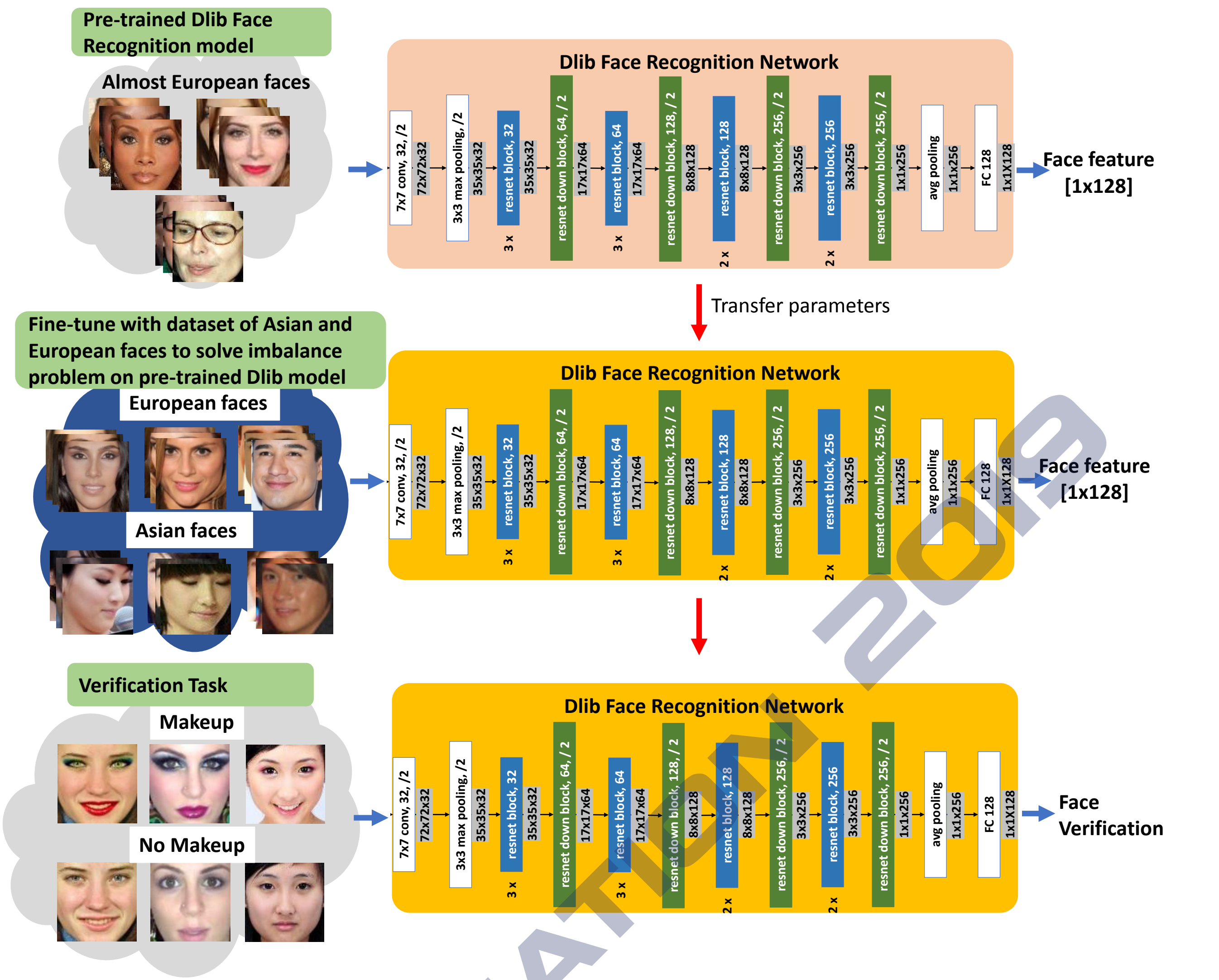


Figure 1. The network structure of pre-trained Dlib face recognition model is shown in orange block. The model was pre-trained on a large dataset of European faces, and is fine-tuned on a balance dataset of Asian and European faces. The fine-tuned model is adopted for makeup face verification task.

BLAN-1 [4] designs a generator and two discriminators to remove makeup, and then extracts features by a Light-CNN pre-trained on MS-Celeb-1M [11]. The similarity metric used in all experiments is cosine distance. However, the de-makeup image is blur. The advantages of style transfer learning based method is that the model can remove the makeup from the face. Nevertheless, the model may remove some inherent features of the faces which leads to blurred and deformed images. The inherent features of the face need to be retained as a specific feature for the face verification process.

A recently proposed BeautyGlow [12] uses Glow [13] to transform image space to latent space. BeautyGlow learns a transformation matrix to extract facial and makeup featrues in latent space. It adjusts makeup features in order to generate light or heavy makeup. However, they do not perform de-makeup task because Glow may not preserve useful information.

B. *Correlation Based Approach*

Correlation-based approach learns the correlation between makeup and non-makeup. It has been known that makeup styles consist of vintage makeup look, Korea makeup look, and cat eyes makeup look [8]. A locality-constrained coupled dictionary learning method learns these styles. It learns pairwise dictionaries (some kinds of transformation from makeup to non-makeup) for three face regions (skin, upper/lower mouth lips, left/right eye) in before and after makeup. Nevertheless, they need to collect more style images.

Correlation mapping methods extract features from local patch and whole facial image [9, 14]. However, the verification accuracy need improvement due to nonlinearity. Therefore, deep learning framework may solve the nonlinear problem due to the use of huge dataset.

C. *Transfer Learning Based Method*

Transfer learning allows us using a pre-trained network on a type of dataset, such as 3 million face images [15], ImageNet [16], and then fine-tune the model on another type of other dataset. The task on pre-trained network can be same as [10, 17] or different as [18]. The study in [18] uses an AlexNet pre-trained on ImageNet-14 for object detection and fine-tuned on CASIA-WebFace [19] for learning multi-pose face. One study extracts features by a pre-trained model and classify patterns by a support vector machine [17]. The study in [10] uses an AlexNet pre-trained on images collected from internet and then fine-tuned on a small makeup dataset [9]. However, the fine-tuned model may not work well once the number of images in the dataset for fine-tune is insufficient, such as makeup face dataset.

The style transfer and learning correlation-based methods require a large pairs of makeup and non-makeup. This may lead to failure when the makeup styles are unseen in training data. This motivates us to apply transfer learning on pre-trained face recognition model, i.e., Dlib [20], which can extract a set of discriminative features from image face and has been shown high performance on LFW dataset. However, such model pre-trained on non-Asian faces may not extract features for Asian faces. To solve this problem, this study proposes a new fine-tune scheme so that the model can extract features for both Asian and European faces and is robust to makeup in face verification.

III. TRANSFER LEARING FOR MAKEUP FACE VERIFICATION

The whole pine-line of proposed method is demonstrated in Figure 1. The proposed framework uses Dlib model [21] as the basis network for transfer learning. The model was pre-trained on a large dataset of European faces, and therefore could not extract features representing Asian faces which usually appear in makeup datasets. A fine-tuned dataset is proposed to adjust the model in order to improve the verification accuracy on both Asian faces and European faces. The proposed framework is implemented by using the network structure and loss definition of the pre-trained model.

A. *Pre-trained Dlib face recognition model*

Dlib [2] is an open source library written by Davis King for computer vision, machine learning, data analysis, optimization and networking [21], [20]. Dlib has been widely applied in both industry and academia domains including robotics and mobile phones. This study uses a pre-trained face recognition Dlib model for face verification. The face recognition model is a ResNet network with 29 convolutional layers [22]. The model was trained on 3 million faces of celebrities including face scrub dataset [23], the VGG dataset [24], and the large number of images crawled from internet. The network structure of pre-trained Dlib model is shown in Figure 1. The feature vector extracted from Dlib model is an embedding vector with 128 dimensions.

B. *Loss function*

The loss function is designed to ensure that a face image of a specific person can be verified as the person. That is the Euclidean distances between all feature vectors of that person should be smaller than a predefined threshold θ. Meanwhile, two face images from two persons cannot be verified as the same person. That is the Euclidean distances between two different persons should be larger than a predefined threshold θ. A negative pair consists of two face images from different persons whose distance is denoted by d_n. A positive pair consists of two face images of the same person whose distance is denoted

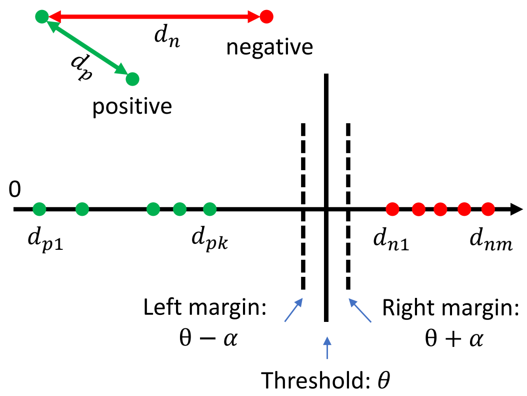


Figure 2 Illustrates conditions for loss function. The figure presented positive distance of k pairs (from d_{p1} to d_{pk}) and negative distance of m pairs (from d_{n1} to d_{nm}).

by d_p. The margin α around decision boundary θ helps distinguish the positive pairs and negative pairs as shown in Figure 2. The threshold and margin in the pre-trained model are designed empirically.

The loss function for negative pair is defined as follows:

$$L_n = \max((\theta + \alpha) - d_n, 0)\,. \tag{1}$$

The loss function for positive pair is defined as follows:

$$L_p = \max(d_p - (\theta - \alpha), 0)\,. \tag{2}$$

The total loss function is defined as follows:

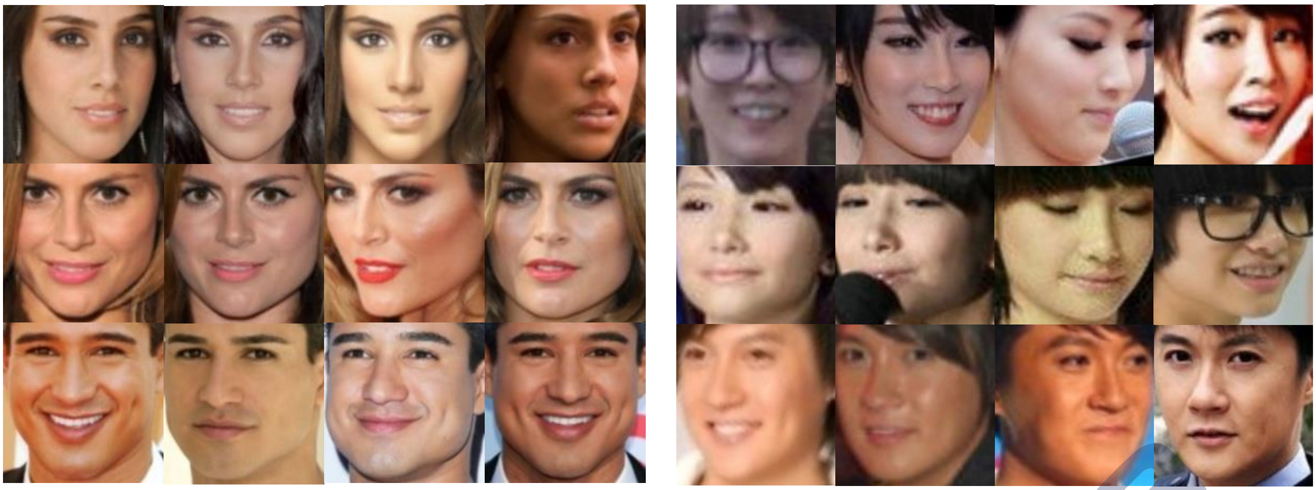
$$L_{total} = L_n + L_p\,. \tag{3}$$

C. *Fine-tune Pipeline*

Dlib is adopted as the pre-trained face recognition model due to its effective feature extraction which is learned from a large dataset of face images and may be robust to makeup. The setting of the learning rate follows the practical advice in [25] where the initial learning rate for fine-tuning should be smaller than learning rate of pre-trained model. The fine-tuning process is shown in Figure 1.

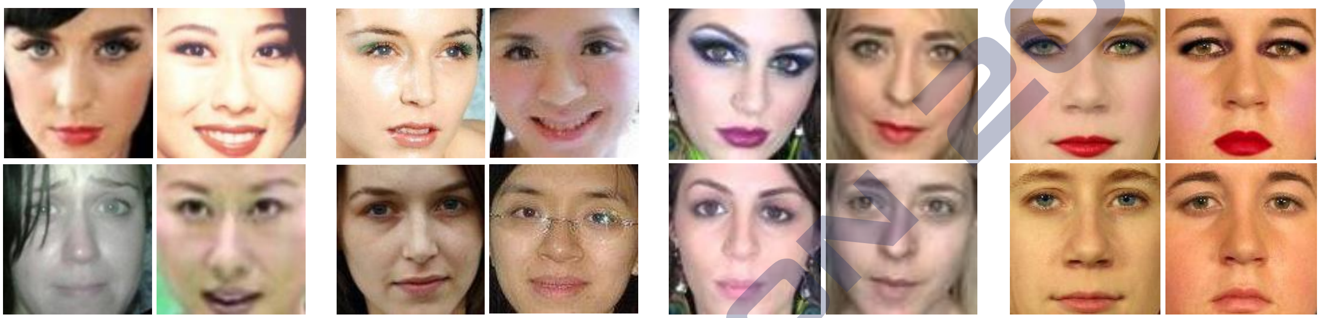
D. *Dataset for fine-tuning*

The purpose of this study is to improve pre-trained Dlib face recognition model so that the model can distinct Asian faces as good as European faces in makeup dataset. A dataset is collected to balance both Asian faces and European faces for fine-tune process. This dataset collects images from Trillion Pairs Challenge [26], Asian-celeb, and Ms-Celeb-1M [11]. Most celebrities in Ms-Celeb-1M are people in America, Great Britain, Albion, Germany, Canada, France, Japan, Italia, Australia, India, and Russian [11] and have been excluded from both LFW [27] and Asian-celeb. Because the number of images for each identity in Trillion-Pairs is different, this study collects 100 images for each identity in order to balance the amount of European faces and Asian faces. Finally, fine-tune dataset consists of 2,482 identities MS-Celeb and 2,482 identities Asian-Celeb, 100 images for each identity. Figure 3 shows few samples of this dataset.



(a) MS-Celeb (b) Asian-Celeb

Figure 3. Some samples of fine-tuning dataset of MS-Celeb and Asian-Celeb.



(a) FAM (b) Dataset1 (c) YMU (d)VMU

Figure 4. Few samples for each dataset. The top row shows makeup faces and the bottom row shows non-makeup faces

E. Evaluation

The performance is evaluated by 5-folds validation [14]. Each image is verified with one positive image and one random selected negative image from different person. A positive pair is a pair of two images from same person, while a negative pair is a pair of two images from two different people. The accuracy in each fold is the average performance of positive and negative pairs. The Euclidean distance is used to determine the similarity. The threshold is selected empirically.

F. Implementation detail

All face images are first detected and cropped using 3D landmark point detection [28]. The face image is resized into 150x150 pixels as the input of Dlib face recognition model. The learning rate is initialized as 0.0001. The learning rate decreases after 5000 iterations when the loss stops decreasing. The fine-tune process stops when the learning rate reach out of 10^{-9}. The threshold and margin value are 0.6 and 0.04, respectively, when computing similarity. Stochastic gradient descent is adopted as optimization algorithm. Each mini-batch in training step contains 20 images of 20 identities. Each image in mini-batch is paired with 3800 positive pairs and 3800 hardest negative pairs selected from 76000 negative pairs according to [20].

The proposed scheme was implemented on CPU Intel® Xeon® CPU E5-2620 v3 @ 2.40GHz -12 threads with GPU NVIDIA 980Ti-7GB and 32GB of RAM. Dataset was loaded on 10 threads. The fine-tune process took 10 hours on the machine.

IV. EXPERIMENTAL RESULT

A. Datasets

The proposed scheme is evaluated on 4 makeup datasets: FAM, Dataset1, YMU, and VMU. The numbers of identities and makeup/non-makeup images in each identity are shown in Table 1. Few samples of the four makeup datasets are shown in Figure 4.

FAM [14]: This dataset contains both male and female face images with different poses and various image qualities. The dataset contains both Asian and European faces with light makeup.

Dataset1 [9]: This dataset consists of Asian and Caucasian face images with frontal view and high image quality. The makeup style is light.

YMU [29]: The makeup in this dataset varies from light to heavy. This dataset includes variations in expression and pose. The illumination condition is constant for the same subject. The hair style before and after makeup drastically varies in few cases [30]

VMU [29]: This dataset contains makeup of lipstick and eye makeup. A full makeup consists of lipstick, foundation, blush and eye makeup. Full makeup faces are used in this study.

(a) (b) (c)

Figure 5. Some false positive pairs in FAM dataset verified by pre-trained dlib model and caused by (a) blurry, (b) face pose, (c) image quality.

Table 1. Four datasets used in this study. The symbol # means the number.

	Year	# identities	# images per identity
FAM [14]	2013	222 males 297 females	1 makeup, 1 non-makeup
Dataset1 [9]	2014	501 females	1 makeup, 1 non-makeup
YMU [29]	2012	151 females	2 before makeup, 2 after makeup
VMU [29]	2012	51 females	1 no makeup, 1 lipstick, 1 eye makeup, 1 full makeover

B. *Evaluation results*

The evaluation result of the proposed transfer learning scheme is compared with FaceNet [1] on the four makeup datasets as shown in Table 3. FaceNet is trained on more than 3 million faces of VGGFace2 dataset [15]. Dlib model could achieve higher accuracy than FaceNet model.

The verification accuracy performed by pre-trained Dlib on YMU and VMU is higher than other methods. However, the verification accuracies are lower than others in FAM and Dataset1. All false positive/negative pairs leading to low accuracy are from FAM and dataset1. For the false positive pairs in FAM, the main reasons are burred images, various poses, and low image quality as shown in Figure 5. Only 3 out of 56 false negative pairs are blurred images. All false negative pairs are Asian faces. The Dataset1 contains only woman faces with high quality image. There are 25 false positive pairs and 41 false negative pairs which are all Asian faces. This indicates the need of fine-tuning Dlib by Asian face images as shown in Table 2.

Table 2 Analysis false positive and negative pairs of European (E) and Asian (A) faces verified by the pre-trained Dlib in FAM [14] and Dataset1 [9] where # A and # E indicate the numbers of Asian and European face images, respectively.

	# A	# E	False Negative		False Positive	
			# A	# E	# A	# E
FAM	419	100	56	0	42	40
Dataset1	328	173	41	0	4	21

In order to show the effect of fine-tuning on Asian faces, the pre-trained Dlib model and fine-tuned model perform feature extraction European faces (473 pairs) and Asian faces (749 pairs) from makeup datasets (FAM, Dataset1, YMU, VMU). Then a distance between the features of a image pair was determined. Figure 6 shows the distance distribution of positive/negative pairs. The threshold values were chosen for each fine-tuned/pre-trained models given minimum number of false pairs on whole makeup datasets. Figure 6 (a) and (b) indicate that the fine-tuned model results in fewer false negative pairs than the pre-trained model for Asian faces. Figure 6 (c) and (d) show that the fine-tuned model results in more false negative pairs but fewer false positive pairs than the pre-trained model for European faces. Figure 6 (a) and (c) show that the pre-trained model results in lower verification accuracy on Asian faces than that on European faces. Figure 6 (b) and (d) shows balanced verification accuracy on both European and Asian faces.

C. *Comparisons with other methods*

The proposed model was compared to other works on four datasets. Table 3 shows that the pre-trained Dlib achieved a very high accuracy on YMU and VMU than the other works, but low accuracy on –FAM when comparing to BLAN2 [5]. The proposed transfer learning scheme helps

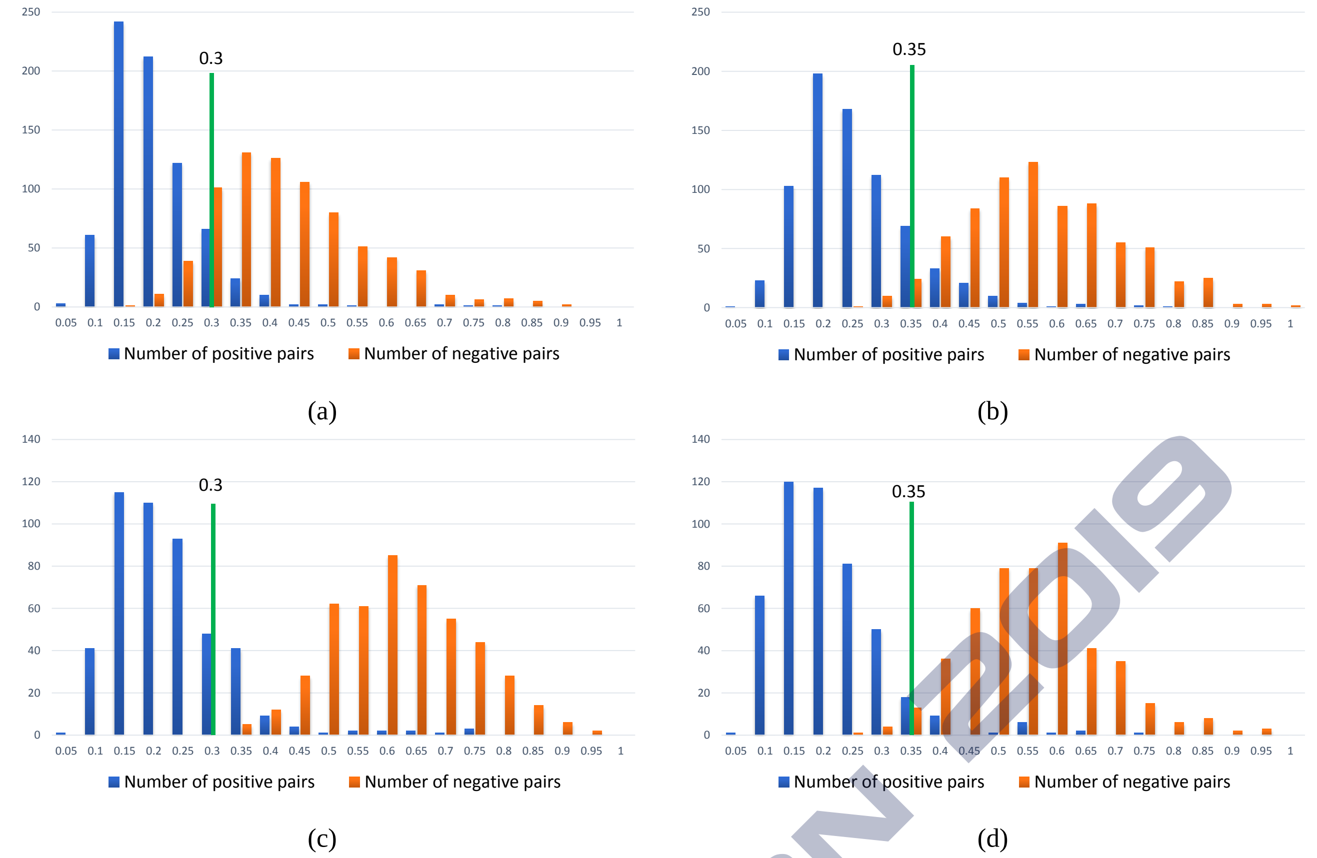


Figure 6 Distance distribution of positive/negative pairs from 749 Asian and 473 European pairs on both pre-trained (two left charts) and fine-tuned (two right charts) Dlib model. The threshold value of 0.3 was chosen for pre-trained model, and 0.35 for fine-tuned model. (a) Distance distribution positive/negative pairs of **Asian** faces determined by **pre-trained** Dlib model (b) Distance distribution positive/negative pairs of **Asian** faces determined by **fine-tuned** Dlib model (c) Distance distribution positive/negative pairs of **European** faces determined by **pre-trained** Dlib model (d) Distance distribution positive/negative pairs of **European** faces determined by **fine-tuned** Dlib model.

the model learn to extract features of both Asian and European faces. The transfer learning scheme make the fine-tuned Dlib model could achieve highest accuracy on all datasets when comparing to other works. The result also shows that the proposed model is robust to the makeup in face verification task.

Table 3. Experimental results where * indicates the model fine-tuned by only Asian face images.

	FAM	Dataset1	YMU	VMU
Facenet [1]	83.60%	90.33%	89.48%	88.05%
BLAN2 [5]	90.00%	95.50%	N/A	N/A
BLAN1 [4]	88.10%	94.80%	N/A	N/A
Muhammad [7]	N/A	N/A	90.04%	92.99%
A.Dantcheva [29]	N/A	N/A	84.11%	95.21%
Pre-trained Dlib	88.90%	92.55%	96.08%	100%
Fine-tuned Dlib*	92.31%	94.45%	85.53%	82.10%
Fine-tuned Dlib	**91.52%**	**96.39%**	**97.37%**	**94.03%**

To show the importance of balance between Asian and European datasets, the Dlib model was fine-tuned by only Asian faces as ablation study. The results show that the accuracy on VMU and YMU are lower than that fine-tuned by both Asian and European faces. This result shows the efficiency of the proposed scheme.

V. Conclusion

This study proposes a transfer learning scheme for makeup face verification by a pre-trained face recognition model, Dlib. A fine-tune dataset is collected from 4,964 identities (100 images/identity) with half of Asian celebrities and half of European celebrities from MS-Celeb and Asian-Celeb. The proposed transfer learning scheme solves the problem of imbalance verification ability on Asian and European face images. The fine-tuned Dlib model could achieve higher verification accuracy on FAM, Dataset1, YMU, and VMU datasets than the other works. Furthermore, this model is robust to makeup, even strong makeup styles from YMU and VMU datasets. Future work will address the problems of low image quality and pose variation.

Acknowledgment

This work was supported in part by the Ministry of Science and Technology (MOST) in Taiwan under grant number of MOST 108-2636-E-006-010 (Young Scholar Fellowship Program). This research was supported in part by the Headquarters of University Advancement at the National Cheng Kung University, which is sponsored by the Ministry of Education, Taiwan. The authors would like to thank Dr. Tyng-Luh Liu for valuable discussion.

References

[1] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 815-823.

[2] D. E. J. J. o. M. L. R. King, "Dlib-ml: A machine learning toolkit," vol. 10, pp. 1755-1758, 2009.

[3] H. Chang, J. Lu, F. Yu, A. J. I. C. C. o. C. V. Finkelstein, and P. Recognition, "PairedCycleGAN: Asymmetric Style Transfer for Applying and Removing Makeup," pp. 40-48, 2018.

[4] Y. Li, L. Song, X. Wu, R. He, and T. Tan, "Anti-Makeup: Learning a bi-level adversarial network for makeup-invariant face verification," in *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018.

[5] Y. Li, L. Song, X. Wu, R. He, and T. J. P. R. Tan, "Learning a bi-level adversarial network with global and local perception for makeup-invariant face verification," vol. 90, pp. 99-108, 2019.

[6] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223-2232.

[7] M. Sajid, N. Ali, S. Hanif Dar, N. I. Ratyal, A. raza Butt, and B. Zafar, *Data Augmentation-Assisted Makeup-Invariant Face Recognition*, 2018.

[8] S. Wang and Y. Fu, *Face Behind Makeup*, 2016.

[9] G. Guo, L. Wen, and S. Yan, "Face Authentication With Makeup Changes," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, pp. 814-825, 2014.

[10] Y. Sun, L. Ren, Z. Wei, B. Liu, Y. Zhai, and S. Liu, *A Weakly-Supervised Method for Makeup-Invariant Face Verification* vol. 66, 2017.

[11] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "MS-Celeb-1M: A Dataset and Benchmark for Large-Scale Face Recognition," *eprint arXiv:1607.08221*, p. arXiv:1607.08221, 2016.

[12] H.-J. Chen, K.-M. Hui, S.-Y. Wang, L.-W. Tsao, H.-H. Shuai, and W.-H. Cheng, "BeautyGlow: On-Demand Makeup Transfer Framework With Reversible Generative Network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 10042-10050.

[13] D. P. Kingma and P. Dhariwal, "Glow: Generative flow with invertible 1x1 convolutions," in *Advances in Neural Information Processing Systems*, 2018, pp. 10215-10224.

[14] J. Hu, Y. Ge, J. Lu, and X. Feng, "Makeup-robust face verification," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 2342-2346.

[15] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018, pp. 67-74.

[16] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, *et al.*, "Imagenet large scale visual recognition challenge," vol. 115, pp. 211-252, 2015.

[17] H. Yu, Z. Luo, and Y. Tang, "Transfer learning for face identification with deep face model," in *2016 7th International Conference on Cloud Computing and Big Data (CCBD)*, 2016, pp. 13-18.

[18] W. AbdAlmageed, Y. Wu, S. Rawls, S. Harel, T. Hassner, I. Masi, *et al.*, "Face recognition using deep multi-pose representations," in *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2016, pp. 1-9.

[19] D. Yi, Z. Lei, S. Liao, and S. Z. J. a. p. a. Li, "Learning face representation from scratch," 2014.

[20] D. King. (2019). *Open source dlib*. Available: https://github.com/davisking/dlib

[21] D. King. (2019, 15-Aug). *Dlib.net*. Available: http://dlib.net/

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[23] H.-W. Ng and S. Winkler, "A data-driven approach to cleaning large face datasets," in *2014 IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 343-347.

[24] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *bmvc*, 2015, p. 6.

[25] S. University. (08-2019). *CS231n Convolutional Neural Networks for Visual Recognition*. Available: http://cs231n.github.io/transfer-learning/

[26] *Trillion Pairs Challenge*. Available: http://trillionpairs.deepglint.com/overview

[27] E. Learned-Miller, G. B. Huang, A. RoyChowdhury, H. Li, and G. Hua, "Labeled faces in the wild: A survey," in *Advances in face detection and facial image analysis*, ed: Springer, 2016, pp. 189-248.

[28] A. Bulat and G. Tzimiropoulos, "How far are we from solving the 2d & 3d face alignment problem?(and a dataset of 230,000 3d facial landmarks)," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 1021-1030.

[29] A. Dantcheva, C. Chen, and A. Ross, "Can facial cosmetics affect the matching accuracy of face recognition systems?," in *2012 IEEE Fifth international conference on biometrics: theory, applications and systems (BTAS)*, 2012, pp. 391-398.