Research Article

Data Augmentation-Assisted Makeup-Invariant Face Recognition

Muhammad Sajid [1], Nouman Ali [2], Saadat Hanif Dar,[2]
Naeem Iqbal Ratyal,[1] Asif Raza Butt,[1] Bushra Zafar,[3] Tamoor Shafique,[4]
Mirza Jabbar Aziz Baig,[5] Imran Riaz,[1] and Shahbaz Baig[5]

[1]*Department of Electrical Engineering, Mirpur University of Science and Technology, Mirpur (A.K.), Pakistan*
[2]*Department of Software Engineering, Mirpur University of Science and Technology, Mirpur (A.K), Pakistan*
[3]*Department of Computer Science, Government College University, Faisalabad, Pakistan*
[4]*Faculty of Computing, Engineering and Science, Staffordshire University, Stoke-on-Trent, UK*
[5]*Department of Electrical (Power) Engineering, Mirpur University of Science and Technology, MUST, Mirpur, Pakistan*

Correspondence should be addressed to Muhammad Sajid; sajid_faiz@hotmail.com

Recently, face datasets containing celebrities photos with facial makeup are growing at exponential rates, making their recognition very challenging. Existing face recognition methods rely on feature extraction and reference reranking to improve the performance. However face images with facial makeup carry inherent ambiguity due to artificial colors, shading, contouring, and varying skin tones, making recognition task more difficult. The problem becomes more confound as the makeup alters the bilateral size and symmetry of the certain face components such as eyes and lips affecting the distinctiveness of faces. The ambiguity becomes even worse when different days bring different facial makeup for celebrities owing to the context of interpersonal situations and current societal makeup trends. To cope with these artificial effects, we propose to use a deep convolutional neural network (dCNN) using augmented face dataset to extract discriminative features from face images containing synthetic makeup variations. The augmented dataset containing original face images and those with synthetic make up variations allows dCNN to learn face features in a variety of facial makeup. We also evaluate the role of partial and full makeup in face images to improve the recognition performance. The experimental results on two challenging face datasets show that the proposed approach can compete with the state of the art.

1. Introduction

With the long-time evolution of social media, collection of face photos of celebrities has shown promising potential to develop face recognition algorithms. The development of such datasets allows the researchers to conduct the research in a variety of ways, such as age invariant face recognition [1–3]. However, subjects in these datasets are public figures and use facial makeup to accentuate their appeal and attractiveness. The resulting cosmetic effects include artificial face colors, shading, contouring, and varying skin tones. Such images differ greatly from the original face images and pose several challenges in recognition tasks [4]. Furthermore, the makeup

wearing increases bilateral size and symmetry of the facial components such as eyes and lips leading to a decreased characteristic distinctiveness of faces [5]. Face recognition systems aim to match query face against gallery faces and this task becomes more difficult in case query face image contains the inherent cosmetic effects which cause absence of discriminative features such as scars, moles, tattoos, microtexture, wrinkles, and pores. For example, some instant skin smoother finishing creams help to reduce the appearance of wrinkles, lines, and pores, altering the quality and color of skin [6]. More precisely, such cosmetic effects can alter the facial shape, perceived size and color of facial parts, location of eyebrows, etc. [4]. Figures 1(a) and 1(b) show an example

FIGURE 1: Example face image (a) before cosmetic application and (b) after cosmetic application (image taken from [9]).

face image before and after application of cosmetic to left side of the face to hide moles.

Keeping in view the substantial ability of makeup to change facial appearance, a number of approaches attempted to achieve makeup-invariant face recognition. Most of the existing methods such as [4, 7, 8] attempt to extract local features from face images to match query faces with gallery. In most of these methods, hand-crafted features like local binary patterns (LBP) Gabor wavelets, and other local and global features. However, these hand-crafted features have two major shortcomings: (i) first, the conventional local features employ hand-crafted mapping, which is not optimal for visual feature representation; (ii) second, some useful information may be compromised in thresholding stage.

To fight the shortcomings of hand-crafted features, deep learning methods have been proposed owing to their ability to learn high level discriminative features automatically. The authors in [11] suggested a weakly supervised method for makeup-invariant face verification. They used a pretrained CNN on Internet videos and fine-tuned small makeup datasets. A voting strategy is then employed to achieve better face verification results across makeup variations. Although CNN are known to be capable of learning the most discriminative features in face images, a frequent problem faced when training CNNs is that there is lack of sufficient data required to maximize the generalization capability of CNNs. This problem becomes even worse in case of makeup-invariant face recognition tasks where existing datasets are quite small. There are many techniques to address this including dropout, transfer learning, and data augmentation. However, not all of these methods will improve the performance, if used intuitively.

The objective of our work is to assess how the discriminative capabilities offered by dCNNs are enhanced when face images with different cosmetic effects are presented to them through an effective data augmentation strategy. We offer the following contributions in this work:

(1) Assess the representation capabilities of dCNNs for makeup wearing face images compared to hand-crafted features

(2) Attempt to enhance face recognition performance using an appropriate data augmentation strategy not reported in the literature before

(3) Determine the individual impact of two distinct data augmentation strategies, namely, the celebrity-famous makeup styles and semantic preserving transformations

(4) We compare the results of our study with existing methods to show the efficacy of the proposed method

The remainder of this paper is organized as follows. Section 2 overviews the existing approaches to makeup-invariant face recognition. The proposed method is presented in Section 3. Experiments and results are presented in Section 4, while results related analysis is presented in Section 5. The last section concludes the study along with future research directions.

2. Related Works

The existing methods to makeup-invariant face recognition can be grouped into two categories with respect to features they used to represent face images, as described in the following subsections.

2.1. Hand-Crafted Feature Based Approaches. Hand-crafted feature based approaches extract local or global features from query and gallery faces to perform matching. The representative works in this category include [4, 7, 8, 12–14]. In [4], authors extracted LBP and Gabor features to capture micropatterns and shape and texture information, respectively. Wang and Kumar [7] used LBP with biometric and nonbiometric blocks. Matching is performed using only the biometric blocks to achieve superior accuracies. In [8], support vector machine (SVM) classifier is trained on depth and texture features to recognize makeup wearing face images. In [12], Gabor wavelets and LBP descriptors are used as shape and texture descriptors respectively, for automatic facial makeup detection with application in face recognition. Similarly, a cosmetic detection approach is presented in [13] by leveraging a multiscale local-global technique. Guo et al. [14] proposed a makeup-invariant face verification strategy based on the feature correlation between original and makeup wearing faces. Additionally, the hand-crafted feature has used other fields of image processing and computer vision such as [15]. The main limitations of hand-crafted features are their fixed encoding and nonoptimal feature limiting the

recognition accuracies. Due to these inherent limitations, hand-crafted features are getting out of style especially after it was shown that CNNs can automatically learn the best features for a given tasks.

2.2. Deep Learning Based Approaches.

The dCNNs have shown promising performance in face recognition problems such as [2, 3, 11, 16–19]. In [2], external features are injected into fully connected layer of a dCNN to achieve better verification accuracies across large age gaps. A triplet deep network is presented in [11] to verify face images across makeup variations. Coupled autoencoders have been used in [16] for face recognition across aging variations. dCNNs along with LBP descriptors have been used for face verification task in [17]. However, most of these approaches focus on age invariant face recognition. Liu et al. [3] proposed deep learning architecture to model aging process and face verification across aging variations. Sun et al. [18] hybrid CNN-Restricted Boltzman Machine (RBM) for face verification in the wild. Li et al. [19] suggested synthesizing nonmakeup images from makeup ones and then using the synthesized nonmakeup images for further verification task. More recently, PairedCycleGAN have been used in [20] to automatically transfer the makeup style of source face image to the target face image. More specifically, the authors suggested a forward and a backward function to apply and remove an example-based makeup style respectively. A patch-based learning approach has been given in [21]. In this approach, the authors used face image patches to encode a set of feature descriptors. An ensemble learning is then followed using random subspace linear discriminant analysis to make the face recognition robust against cosmetic variations.

Although dCNNs are powerful architectures capable of yielding competitive performance across a variety of classification tasks, yet their performance is limited by the unavailability of enough training data. A prominent example in this regard is makeup wearing face images datasets which are relatively smaller datasets. The existing deep models to makeup-invariant face verification studies mainly rely on transfer learning such [11] as or face synthesis [19] to cope the negative effects of facial cosmetics on verification accuracy. However, we believe that transfer learning and face synthesis approaches are not suitable to fight appearance variations. This is because, in case of transfer learning, it is difficult to extract identity specific features from CNN layers in the presence of cosmetic effects. Similarly, face synthesis results in pseudofaces with inferred content of lower quality, resulting in lower verification accuracies. Therefore, in the presence of different makeup styles, it is more convenient to use makeup style-aware data augmentation to achieve invariance across cosmetic variations. Our idea is to augment the training dataset in a way that produces the competitive results for makeup-invariant face verification task through the intelligent creation of celebrity-famous makeup styles.

2.3. Makeup Styles-Aware Data Augmentation.

The efficacy of dCNNs is known to be strongly dependent on availability of sufficient training dataset. Not having enough diverse and quality data in training will result in overfitting which means that dCNN is highly biased to the training data and will generate errors in the testing stage. To address overfitting, data augmentation is an effective approach which aims to expand the training data by applying transformations to existing samples resulting in new samples as additional training data. It has been shown that intelligent data augmentation during training phase can improve the generalization ability of dCNNs. However, the choice of augmentation strategy can be more important than the type of network architecture used [22]. Many existing methods suggest how to augment datasets while training dCNNs. For example, Simard et al. [23] suggested different augmentation techniques suitable for variety of classification tasks. Similarly, [24] suggests random flip, rotate, and scale face images to achieve data augmentation. In [25], the authors leverage neural networks to learn optimal data augmentation suitable for image classification. The above methods suggest that data augmentation is performed according to the desired task. Like the above augmentation methods, the proposed makeup styles-aware data augmentation attempts to address the issue of limiting training data in facial makeup datasets to fight overfitting and improve regularization. In this work, we propose to use a more advanced method to augment the training data by makeup transformations aimed at allowing dCNNs to learn features robust to different cosmetic variations. More precisely, training data is augmented by developing different makeup styles for a given face image in the training dataset. This will allow dCNNs to learn features across a variety of makeup styles. Since the subject-specific cosmetic variations can change significantly according to the occasions, we propose to develop six different celebrity-famous makeup styles [9, 26], namely, (i) *Everyday*: it is a casual light facial enhancement makeup for daily use; (ii) *Professional*: it is an event-specific makeup wearing; (iii) *Smoky*: it is usually suggested for night time events and parties; (iv) *Asian*: it involves alluring and sultry eye and lips makeup; (v) *Disco*: it includes light blue eyeshadow and bold eyelashes for enhancement; (vi) *Gothic*: the gothic look is a dark makeup style and a favorite among teenagers. In addition to six different makeup styles, we also augment data by using detexturized, decolorized and edge map of each original face image. In this way, we develop total 9 variants for each original face image in the augmented dataset. The makeup style images will allow the dCNN to learn variety of features across face images wearing cosmetics. Detexturized face images will allow dCNN to learn discriminative features in the presence of foundation makeup used to cover facial flaws, such as moles, scars, and open pores. Decolorized face images will allow them to be matched with the images wearing black-and-white *Halloween* makeup. Finally, the edge map of a given face image will enable the dCNN to learn facial lines to be matched with make wearing images with contouring. Contouring involves giving certain shape to facial components such as nose bridge and lips [6].

The six makeup styles are implemented using a publicly available makeup synthesis tool called TAAZ [10]. The TAAZ facilitates the application of symmetric makeup to a given face image. The available options to apply synthetic makeup include various accessories, complete looks, Halloween, etc. For example, the tool allows applying four unique symmetric

FIGURE 2: Four unique symmetric blush patterns used in TAAZ (image taken from [10]).
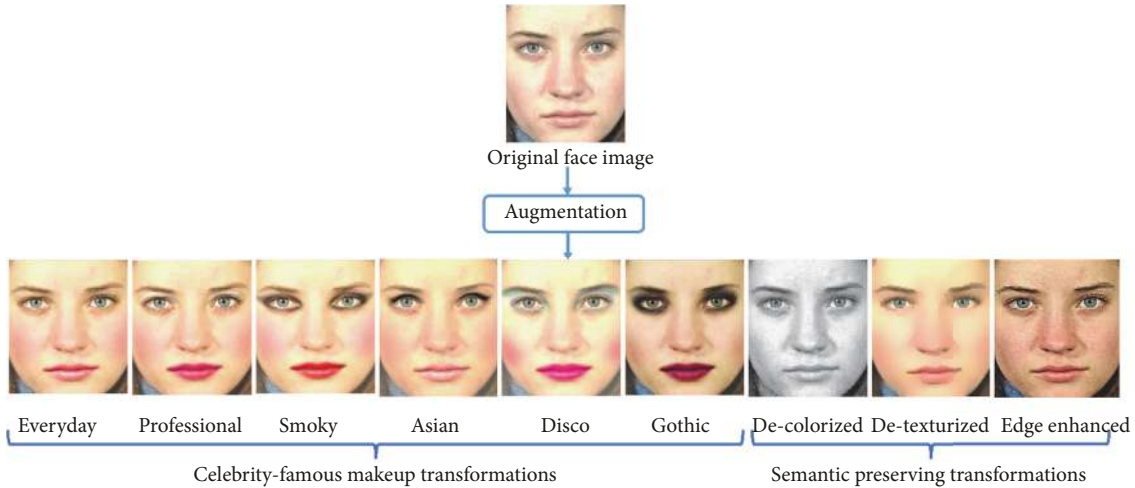

FIGURE 3: Data augmentation using original face image and 6 celebrity-famous makeup and 3 semantic preserving transformations.

blush patterns to a given face image with different options shown in Figure 2. The tool box also allows choosing among three different levels including light, medium, and dark for foundation and concealer.

The decolorized face images are obtained by transforming color face images into gray scale by using a luminance model adopted by NTSC and JPEG [27]. In a similar fashion, detexturized images are obtained by using anisotropic diffusion approach [28], which aims to remove texture by preserving edges and smooth homogeneous regions. We use unsharp masking for edge enhancement. The augmented set for a given original face image is shown in Figure 3.

3. Deep Makeup-Invariant Face Verification

The proposed makeup-invariant face verification method consists of the following steps: (i) preprocessing; (ii) data augmentation; (iii) face verification. The details of each step are given in the following subsections.

3.1. Preprocessing. To alleviate the unwanted illumination and translational variations, the face images are preprocessed as described below.

(1) Face images are aligned according to the centers of two such that they are vertically upright.

(2) To eliminate the illumination variations, Difference of Gaussian (DoG) filtering has been used [29].

(3) Finally, the face images are cropped and resized to 200x200 pixels size.

3.2. Data Augmentation. During the training phase, we use dCNNs to perform face verification task. Our objective is to train the dCNNs for the specified task while there are not enough representative face images in the given dataset. To do so, we use data augmentation suitable for makeup-invariant face recognition task. Training data is augmented by developing six celebrity-famous makeup styles for each original face image using state-of-the-art TAAZ application. Additionally, three semantic preserving variants are developed. Recently, such variants have been effectively utilized in deep learning based visual search [30]. It is worthwhile to mention that labels remain unchanged after applying the makeup style or semantic preserving transformations. The only constraint of the proposed data augmentation is that each face image should follow the suggested makeup and semantic preserving transformations. In the final dataset, each original face image has 9 other versions which sufficiently enlarge the dataset suitable for training the dCNNs.

3.3. Face Verification. Face verification aims to distinguish whether a pair of face images has the same identity [31]. Recently, there is growing interest and significant progress in face verification owing to various applications [3, 17, 18]. In this work, we propose to use dCNN model to learn the discriminative face features in the presence of facial makeup variations. Using the facial features of each face image as ground-truth, the dCNN is trained using augmented dataset. Given a makeup wearing image as input, a parallel dCNN will extract the discriminative features to decide the identity
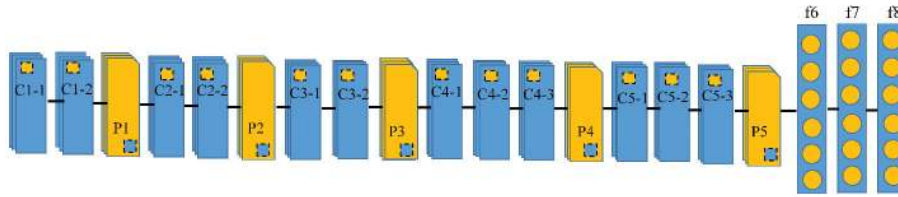
FIGURE 4: The architecture of the VGGNet.

of the test image. To this end, we first review some basic concepts of the proposed dCNN model and then go ahead to our proposed face verification methodology.

Recently, dCNNs have emerged as powerful architectures capable of learning discriminative features automatically. They have shown competitive performance in recognizing face images across different challenges such as aging, pose, and makeup [3, 11, 16, 32]. However, their application to existing datasets is limited by overfitting owing to small number of face images. To handle this challenge, data augmentation is an effective approach. A typical dCNN is a hierarchical architecture composed of different data processing layers such that output of a layer becomes the input of following layer and so on. The most prominent layers include the convolutional layers and pooling layers. The convolutional layers apply learned kernels to the input data and generate feature maps representing the discriminative features. The convolutional layers are followed by the pooling layers which aim to extract the most abstract information from the underlying feature maps. Often the dCNN ends with fully connected layers which model the higher level features.

In this study, we propose to use visual geometry group (VGG) deep network architecture [33]. The choice of network is motivated by its better performance for the selected makeup-invariant face recognition task. We use the variant called VGG-19 containing 16 convolutional (C), 5 pooling (P) and 3 fully connected (f) layers. Each stack of convolutional layers is followed by a pooling layer. The depths of first three stacks of convolutional layers are 64, 128, and 256 respectively, while last two convolutional stacks have depth of 512 each. The size of each fully connected layer is 4096. Finally, the softmax layer produces the similarity score for a given pair of face images. The VGGNet uses filter size of 3x3. The combination of two 3x3 filters results in a receptive field of 5x5 retaining the benefits of smaller filters. The network is famous for its deeper yet simpler architecture as shown in Figure 4.

We train the chosen dCNN model on augmented dataset for face verification task, where our purpose is to predict whether a pair of face images belongs to same person or not.

During the training stage, our goal is to train the dCNN model to perform the desired face verification task while there are not enough training samples in the given dataset. To solve this problem, we use the suggested data augmentation. For a given original face image in the training dataset, we develop 6 celebrity-famous makeup styles and 3 semantic preserving transformations as shown in Figure 3. Such data augmentation has the advantage that the dCNN can learn the possible cosmetic variations caused by wearing facial makeup. The dCNN model used in this study has been trained using stochastic gradient decent (SGD) and standard backward propagation [34].

4. Experiments and Results

We evaluate our data augmentation-assisted makeup-invariant approach on 2 standard datasets including YouTube Makeup Database (YMU) Database [4], and Virtual Makeup (VMU) Database [4]. The YMU dataset contains the original and makeup wearing face images of 99 Caucasian women. The dataset is collected from YouTube video tutorials. The VMU dataset contains face images of 51 Caucasians before and after applying synthetic makeup. The dataset contains three synthetic makeup styles including only lipstick wearing, eye makeup and full makeup. Figure 5 shows example face image pairs from these datasets.

4.1. Experiments on the YMU and VMU Datasets. In this subsection, we present 5 distinct techniques to conduct face recognition experiments on YMU and VMU datasets.

4.1.1. Experiment 1. In this experiment, we examine the suitability of the proposed data augmentation strategy to verify original face images against natural makeup wearing face images. To this end, we train VGGNet from scratch for face verification task. In training stage, the initial learning rate is set at 0.001 which is decreased by a factor of 10 after every 100 epochs. For each original face images we develop 6 celebrity-famous makeup styles and 3 semantic preserving transformation. In this way, we develop a total 9900 face images for 99 original face images in the augmented dataset. Similarly, we generate an augmented dataset for VMU dataset resulting in 5500 face images for 55 original subjects. We apply 200x200 RGB face images as input to the network which outputs the similarity of a pair of face images using a binary softmax classifier. More precisely, we train the dCNN model using augmented YMU dataset. The makeup wearing face images from both datasets are used as test sets to judge the efficacy of the proposed approach. The test accuracies for this series of experiments are shown in Table 1. We use the results of this experiment to create a baseline against which results from other methods can be compared.

For the same experimental setup, we report the test accuracies for the following experiments.

4.1.2. Experiment 2. In this series of experiments, we train the dCNN model without data augmentation. We use original
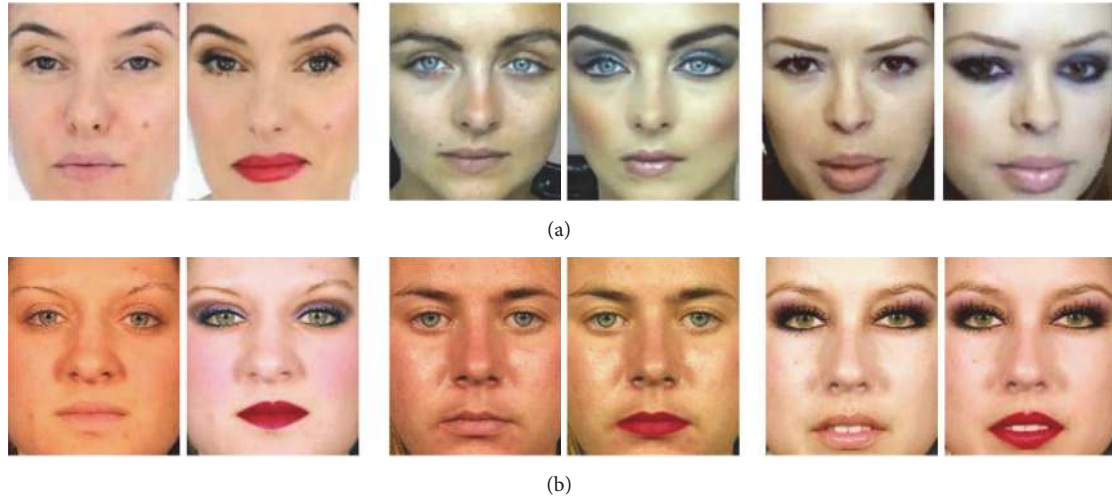
(a)

(b)

FIGURE 5: Example face image pairs from (a) YMU; (b) VMU dataset. In each pair, image on left side is original while right sided image is wearing facial makeup.

TABLE 1: Comparison of the proposed techniques.

Experiment No.	Augmentation	Dataset	Test accuracy
1	6 celebrity-famous makeup styles + 3 semantic preserving transformation (the proposed augmentation-assisted approach)	YMU	90.04
		VMU	92.99
2	No data augmentation is used	YMU	56.00
		VMU	57.85
3	6 celebrity-famous makeup styles	YMU	75.61
		VMU	78.19
4	3 semantic-preserving transformations	YMU	67.66
		VMU	68.05
5	Transfer learning without data augmentation	YMU	86.37
		VMU	88.48

face from YMU and VMU datasets in training while makeup wearing face images are used in testing. The test results are shown in Table 1.

4.1.3. Experiment 3. In the third series of experiments, we train the dCNN model with data augmentation using celebrity-famous makeup styles only. To this end, we develop an augmented dataset consisting of 693 face images for YMU dataset and 350 face images for VMU dataset using the chosen celebrity-famous makeup styles. Makeup wearing face images are then tested with results reported in Table 1.

4.1.4. Experiment 4. In this series of experiments, the chosen dCNN model is trained on data augmentation using semantic preserving transformations only. For this purpose we expand the training datasets of YMU and VMU datasets using three sematic preserving methods for each original face image. This results in 396 and 200 face images for YMU and VMU training sets respectively. The facial makeup wearing face images are presented to the VGGNet for testing. The test results are reported in Table 1.

4.1.5. Experiment 5. In the last series of experiments, we pretrain the VGGNet on a large dataset called the Cross Age Celebrity Dataset (CACD) [1] and fine tune on makeup datasets including YMU and VMU. The motivation behind this series of experiments is to answer the question if transfer learning is helpful in recognizing makeup wearing face images. In transfer learning, we learn the knowledge by training the dCNN on some large dataset to avoid overfitting. The learned knowledge is then used to solve recognition task on small datasets more effectively. In our case, the large dataset is CACD, while smaller datasets are YMU and VMU. It is worthwhile to note that the CACD dataset have been collected for 160000 face images of more than 2000 celebrities. Since the celebrities use facial makeup frequently, therefore we want to know whether transfer learning via fine tuning can be helpful in recognizing face images of the celebrity photos present in the small datasets including YMU and VMU. When fine tuning, the last fully connected layer of VGGNet is replaced by a binary softmax classifier. Face verification experiments are then conducted on facial

TABLE 2: Comparison of the EER for the proposed and existing methods.

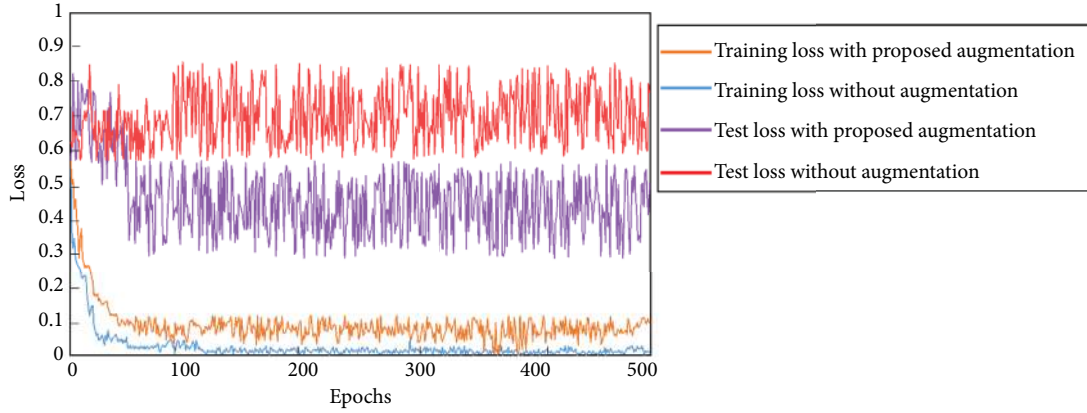| Approach	EER (%)	
	YMU	VMU
Local Gabor Binary Patterns [4]	15.89	5.42
Data augmentation using GAN [35]	11.24	4.99
Proposed data augmented-assisted approach	9.61	2.97

FIGURE 6: Training and test losses for experiments 1 and 2 on YMU dataset.

makeup wearing face images collected from YMU and VMU test sets. The test results for this series of experiments are presented in Table 1.

4.2. Comparison with Existing Methods. We also compare the results of the proposed approach with the closest competitor approach presented in [4]. We calculate and compare the equal error rates (EER) for the proposed approach with those presented in the existing study. It is worthwhile to mention that equal EER is a measure to evaluate the performance of face recognition algorithms. Since the choice of augmentation strategy directly affects the matching accuracy of face images, we also compare the EER of the proposed approach data augmentation approach with generative adversarial network (GAN) [35] based data augmentation. More precisely, the GAN is used to synthesize face images with suggested makeup variations. Classically, the GAN structure consists of a generator and a discriminator in an adversarial environment. The discriminator is used to discern the samples from model and training data. In contrast, the generator aims to maximally confuse the discriminator. For each original face image, 6 makeup wearing face images are generated using GANs. The generated face images are then used to train the CNN for face recognition task. Table 2 compares the corresponding results for the experimental protocol given in [4] when original face images are matched with makeup wearing photos, both for YMU and VMU datasets.

5. Analysis

In this section, we give an analysis of the results presented in the previous section as follows:

(i) The impact of proposed data augmentation strategy is analyzed by tracing the training and test losses for 500 epochs for both the YMU and VMU datasets as shown in Figures 6 and 7 respectively. It is obvious that the overfitting becomes reduced when the proposed data augmentation strategy is applied. In contrast, there is massive overfitting when no data augmentation was applied to train the dCNN. The smaller difference between training and test losses caused by the proposed augmentation method shows how this strategy is helpful for the dCNN to learn the most discriminative features for the desired task. More precisely, the dCNN learns the celebrity-famous makeup styles and semantic preserving features in training stage. In testing stage, a face image wearing an arbitrary makeup can be easily recognized based on this aggressive training. This suggests the effectiveness of the method to prevent the dCNN from overfitting and gives superior performance to recognize face images wearing makeup.

(ii) We observe an improvement in accuracy from 56.00% to 90.04% when proposed augmentation is employed in case of YMU dataset. In case of VMU dataset an improvement from 57.85 to 92.99% is observed. These results show a significant improvement in test accuracy which can be attributed to the task-specific features learned by the dCNN using the proposed augmentation strategy. The cause of poor accuracies of the experiments without data augmentation is the massive overfitting caused by smaller datasets.

(iii) From experimental results reported in Table 1, we can see that, among the proposed data augmentation techniques, the combination of celebrity-famous makeup styles and semantic preserving transformations in data augmentation exhibits the highest test accuracies. In contrast, the individual data augmentation is using only the celebrity-famous
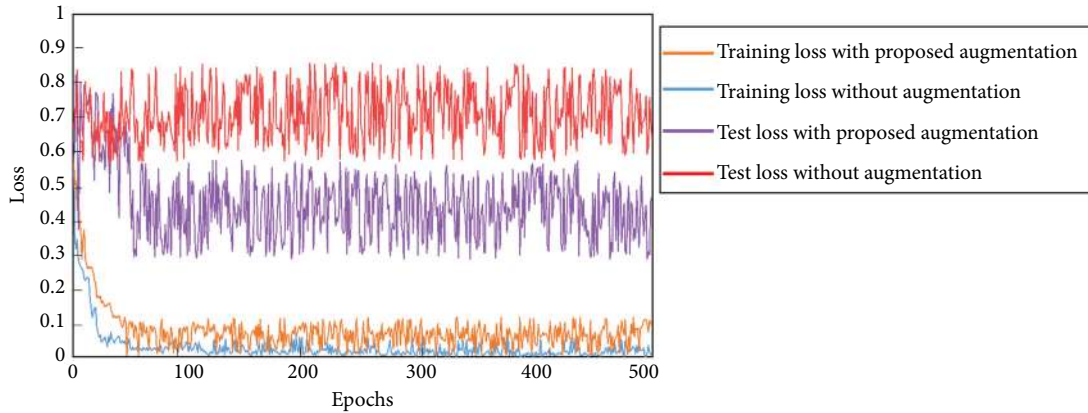
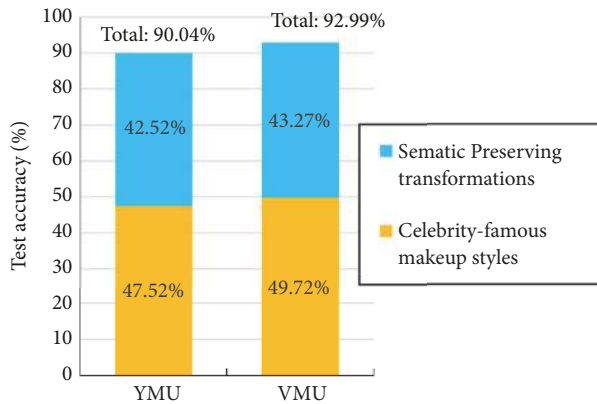FIGURE 7: Training and test losses for experiments 1 and 2 on VMU dataset.

FIGURE 8: Part-to-whole comparison of test accuracies for chosen augmentation strategies for YMU and VMU datasets.

makeup styles or semantic preserving transformations. This is because, semantic preserving transformations adds additional cosmetic related information to the augmented dataset. More precisely, among three semantic transformations, the decolorized images compensate for the Charcoal and gray makeup [9]. The detexturized face images compensate for the application of foundation makeup only not covered under 6 celebrity-famous makeup styles [4]. Finally, the edge-enhanced face images can compensate for facial countering used by celebrities to shape the facial components like nose bridge and lips [9], thus combining celebrity-famous makeup styles with semantic preserving transformations in data augmentation results in superior accuracies.

(iv) Figure 8 shows the part-to-whole comparisons of the percentage accuracies achieved by the celebrity-famous makeup styles and semantic preserving transformations for the proposed augmented-assisted makeup-invariant face recognition task. One can observe the dominant role of celebrity-famous makeup style compared to the semantic preserving transformations in recognizing face images wearing makeups. This is because the celebrity-famous makeup styles are more frequently followed by the subjects present in the makeup datasets. In contrast, comparatively lesser number of

subjects follow the semantic preserving makeup transformations. This also suggests the popularity of celebrity-famous makeup styles among the celebrities compared to the later makeup variant.

(v) From the comparisons shown in Table 1, it is obvious that our method surpasses the accuracies achieved by the transfer learning strategy. Compared to the accuracies of 86.37% on YMU and 88.48% on VMU datasets for transfer learning approach, the proposed method achieves test accuracies of 90.04% and 92.99% on YMU and VMU datasets respectively. In transfer learning the knowledge learned on a large dataset (CACD in this study) is used to solve the new problem (recognizing face images from YMU and VMU datasets) effectively. Recall that for transfer learning experiment, we used VGGNet pretrained on CACD dataset. Although the CACD dataset contains celebrities face images but does not necessarily contains face images with semantic transformations. This results in poor feature learning and subsequent discriminative knowledge transfer when VGGNet is fine-tuned for smaller datasets including YMU and VMU. This suggests the superiority of the proposed data augmentation strategy over transfer learning approach for the desired task of makeup-invariant face recognition.

(vi) Comparison of the test accuracies on two selected datasets suggests better results for VMU compared to YMU dataset across all the reported techniques in our study. This is because YMU is more challenging dataset containing celebrity photos with real makeups. In contrast, VMU contains original photos with synthetic makeups with only three makeup variations.

(vii) The comparative results presented in this study suggest that the proposed approach is effective in recognizing face images across a variety of facial makeup variations including both the real and synthetic makeup variations.

(viii) In Table 2, we compared the EER of the proposed method with an existing method for similar experimental setup. One can observe that the proposed approach exhibits fairly low error rates on both the YMU and VMU dataset. The superior performance of the proposed approach can be attributed to: (i) the effective data augmentation strategy suitable for recognizing makeup wearing face images, and

(ii) the automatically learned deep face features compared to the hand-crafted features employed in [4]. The proposed approach also gives superior performance compared the augmentation strategy based on GANs. This is because, face images generated using GANs are pseudofaces of lower quality containing the inferred content. The pseudofaces result in poor matching accuracy due to poor quality [36]. This results in higher EER for GAN based generated face images. In contrast, the proposed approach uses the original face images of high quality without any inferred content.

6. Conclusions

In this paper we presented a new data augmentation-assisted makeup-invariant face recognition approach. The proposed approach has shown promising results on two standard datasets. Particularly we discussed data augmentation approach based on celebrity-famous makeup styles and semantic preserving transformations suitable for makeup-invariant face recognition. We focused on training a dCNN model to effectively learn the discriminative features in the presence of cosmetic variations and to fight the frequently occurring problem of overfitting in dCNNs. The different approaches presented in this study suggest that two presented data augmentation strategies together can yield superior test accuracies. The experimental results on YMU and VMU datasets suggest that the makeup-aware data augmentation is better than the transfer learning through fine tuning a dCNN for the desired task. The experimental results suggest that celebrity-famous makeup styles is frequently followed by the subjects present in the chosen datasets. Finally, the comparative results show that the proposed approach can compete with the existing methods well.

Future studies may include more sophisticated data augmentation learning strategies to fight both the overfitting and makeup variations.

Data Availability

Previously reported face image datasets including the YouTube Makeup Database (YMU) Database and Virtual Makeup (VMU) Database have been used to support this study. The datasets are available upon request from the sponsors. The related studies using these datasets are cited at relevant places within the text as [4].

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

Muhammad Sajid and Tamoor Shafique conceived the idea. Muhammad Sajid, Mirza Jabbar Aziz Baig, and Imran Riaz conducted the experiments and wrote the manuscript. Muhammad Sajid and Shabaz Baig made the tables and figures. Nouman Ali, Saadat Hanif Dar, Naeem Iqbal Ratyal, Asif Raza Butt, and Bushra Zafar took part in revising the manuscript with contribution in conducting the relevant

experiments. All authors contributed to the final preparation of the manuscript.

References

[1] B.-C. Chen, C.-S. Chen, and W. H. Hsu, "Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset," *IEEE Transactions on Multimedia*, vol. 17, no. 6, pp. 804–815, 2015.

[2] S. Bianco, "Large Age-Gap face verification by feature injection in deep networks," *Pattern Recognition Letters*, vol. 90, pp. 36–42, 2017.

[3] L. Liu, C. Xiong, H. Zhang, Z. Niu, M. Wang, and S. Yan, "Deep Aging Face Verification With Large Gaps," *IEEE Transactions on Multimedia*, vol. 18, no. 1, pp. 64–75, 2016.

[4] A. Dantcheva, C. Chen, and A. Ross, "Can facial cosmetics affect the matching accuracy of face recognition systems?" in *Proceedings of the 5th IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS '12)*, pp. 391–398, Arlington, Va, USA, September 2012.

[5] G. Rhodes, A. Sumich, and G. Byatt, "Are average facial configurations attractive only because of their symmetry?" *Psychological Science*, vol. 10, no. 1, pp. 52–58, 1999.

[6] "L'ORE'L PARIS," https://www.lorealparisusa.com/beauty-magazine/makeup/face-makeup/how-to-hide-large-pores.aspx, [Accessed 28 Decebmer 2017].

[7] T. Y. Wang and A. Kumar, "Recognizing human faces under disguise and makeup," in *Proceedings of the 2nd IEEE International Conference on Identity, Security and Behavior Analysis, ISBA 2016*, Japan, March 2016.

[8] A. Moeini, K. Faez, and H. Moeini, "Face recognition across makeup and plastic surgery from real-world images," *Journal of Electronic Imaging*, vol. 24, no. 5, 2015.

[9] M.-L. Eckert, N. Kose, and J.-L. Dugelay, "Facial cosmetics database and impact analysis on automatic face recognition," in *Proceedings of the 2013 IEEE 15th International Workshop on Multimedia Signal Processing, MMSP 2013*, pp. 434–439, Italy, October 2013.

[10] http://www.taaz.com/.

[11] Y. Sun, L. Ren, Z. Wei, B. Liu, Y. Zhai, and S. Liu, "A weakly supervised method for makeup-invariant face verification," *Pattern Recognition*, vol. 66, pp. 153–159, 2017.

[12] C. Chen, A. Dantcheva, and A. Ross, "Automatic facial makeup detection with application in face recognition," in *Proceedings of the 6th IAPR International Conference on Biometrics, ICB 2013*, Spain, June 2013.

[13] O. Sharifi and M. Eskandari, "Cosmetic Detection Framework for Face and Iris Biometrics," *Symmetry*, vol. 10, no. 4, p. 122, 2018.

[14] G. Guo, L. Wen, and S. Yan, "Face authentication with makeup changes," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 5, pp. 814–825, 2014.

[15] H. Yang and L. Yu, "Feature extraction of wood-hole defects using wavelet-based ultrasonic testing," *Journal of Forestry Research*, vol. 28, no. 2, pp. 395–402, 2017.

[16] C. Xu, Q. Liu, and M. Ye, "Age invariant face recognition and retrieval by coupled auto-encoder networks," *Neurocomputing*, vol. 222, pp. 62–71, 2017.

[17] H. Zhai, C. Liu, H. Dong, Y. Ji, Y. Guo, and S. Gong, "Face Verification Across Aging Based on Deep Convolutional Networks and Local Binary Patterns," in *Proceedings of the International Conference on Intelligent Science and Big Data Engineering*, 2015.

[18] Y. Sun and X. Wang, "Hybrid deep learning for face verification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 10, 2016.

[19] Y. Li, L. Song, X. Wu, R. He, and T. Tan, "Anti-makeup: Learning bi-level adversarial network for makeup-invariant face verification," https://arxiv.org/abs/1709.03654, 2017.

[20] H. Chang, J. Lu, F. Yu, and A. Finkelstein, "PairedCycle-GAN: Asymmetric Style Transfer for Applying and Removing Makeup," in *Proceedings of the CVPR 2018*, 2018.

[21] C. Chen, A. Dantcheva, and A. Ross, "An ensemble of patch-based subspaces for makeup-robust face recognition," *Information Fusion*, vol. 32, pp. 80–92, 2015.

[22] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, Cambridge, Mass, USA, 2016.

[23] P. Y. Simard, D. Steinkraus, and J. C. Platt, "Best practices for convolutional neural networks applied to visual document analysis," in *Proceedings of the 7th International Conference on Document Analysis and Recognition (ICDAR '03)*, pp. 958–963, August 2003.

[24] F. Chollet, "Keras," https://github.com/, 2015.

[25] J. Lemley, S. Bazrafkan, and P. Corcoran, "Smart Augmentation Learning an Optimal Data Augmentation Strategy," *IEEE Access*, vol. 5, pp. 5858–5869, 2017.

[26] T. Alashkar, S. Jiang, and Y. Fu, "Rule-based facial makeup recommendation system," in *Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition*, 2017.

[27] H. Han, C. Otto, X. Liu, and A. K. Jain, "Demographic estimation from face images: Human vs. machine performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 6, pp. 1148–1161, 2015.

[28] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

[29] H. Han, S. Shan, X. Chen, and W. Gao, "A comparative study on illumination preprocessing in face recognition," *Pattern Recognition*, vol. 46, no. 6, pp. 1691–1699, 2013.

[30] J. Ahmad, K. Muhammad, S. W. Baik, and Z. Lv, "Data augmentation-assisted deep learning of hand-drawn partially colored sketches for visual search," *PLoS ONE*, vol. 12, no. 8, Article ID e0183838, 2017.

[31] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, *Labeled Faces in The Wild: A Database for Studying Face Recognition in Uncostrained Environments*, University of Massachusetts, Amherst, MA, USA, 2007.

[32] I. Masi, F.-J. Chang, J. Choi et al., "Learning Pose-Aware Models for Pose-Invariant Face Recognition in the Wild," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.

[33] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proceedings of the International Conference on Learning Representations*, 2015.

[34] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.

[35] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza et al., "Generative adversarial nets," *Advances in Neural Information Processing Systems*, vol. 2, pp. 2672–2680, 2014.

[36] C. Otto, H. Han, and A. K. Jain, "How does aging affect facial components," in *Proceedings of the 12th International Conference on Computer Vision*, 2012.