Detecting Facial Retouching Using Supervised Deep Learning

Aparna Bharati, Richa Singh, *Senior Member, IEEE,* Mayank Vatsa, *Senior Member, IEEE,* Kevin W. Bowyer, *Fellow, IEEE*

***Abstract*—Digitally altering, or "retouching", face images is a common practice for images on social media, photo sharing websites, and even identification cards when the standards are not strictly enforced. This research demonstrates the effect of digital alterations on the performance of automatic face recognition, and also introduces an algorithm to classify face images as original or retouched with good high performance. We first introduce two face image databases with unaltered and retouched images. Face recognition experiments performed on these databases show that when a retouched image is matched with its original image or an unaltered gallery image, the identification performance is considerably degraded, with a drop in matching accuracy of up to 25%. However, when images are retouched with the same style, the matching accuracy can be misleadingly high in comparison to matching original images. To detect retouching in face images, a novel supervised deep Boltzmann machine algorithm is proposed. It uses facial parts to learn discriminative features to classify face images as original or retouched. The proposed approach for classifying images as original or retouched yields an accuracy of over 87% on the datasets introduced in this paper and over 99% on three other makeup datasets used by previous researchers. This is a substantial increase in accuracy over the previous state-of-the-art algorithm [5] which has shown less than 50% accuracy in classifying original and retouched images from the ND-IIITD Retouched Faces database.**

Fig. 1. Illustration of the effect of retouching face images. In each pair of images, one is original and one is retouched. Images in first two rows are obtained from internet.

I. Introduction

Face recognition is being increasingly used for both personal and security applications. Several of these applications such as controlled user authentication require a human in the loop. However, unattended applications such as surveillance, auto-tagging in media collection, and law enforcement require handling several other covariates such as disguise, aging, plastic surgery, and low resolution. Another covariate, which has received very little attention to date in the biometrics literature, is matching photographic images with *retouched* (tampered/doctored) face images. This covariate is particularly pertinent in the following scenarios:

- In digital media collections such as Flickr, Picasa, and Facebook, the images are often *retouched* by the users (for beautification purposes) before uploading. The retouching process includes altering facial features in various ways: "airbrush out" pimples, age spots and wrinkles, make the whites of the eyes whiter, make the teeth whiter, change shape of nose and eyebrows, remove wrinkles, add texture, adjust skin tone, and make the face slimmer or fuller. If these images are used for auto-tagging, the face recognition algorithm may not yield correct results.
- In several photo-identification documents such as driver's license, passport, and other government issued id-cards, hard-copy photographs are required. In general, the user goes to a photo studio, gets the image captured, gets it retouched for better print quality, and uses the print of this image for application. Due to lack of process where ISO standards are strictly followed, these images are used for creating identification documents including passport. Then this retouched image may serve as an enrollment image and be matched with real-time photographic (unretouched) images.
- Digital retouching can be used to alter the appearance of models in the fashion industry. Such "skinny" appearances can distort the mindset of the general population towards weight loss and cause anorexia or other disorders. As mentioned in [1], "*Media-portrayed idealized images*

A. Bharati, R. Singh, and M. Vatsa are with the Indraprastha Institute of Information Technology (IIIT) Delhi, India, e-mail: {aparna11027, rsingh, mayank}@iiitd.ac.in. K. W. Bowyer is with the University of Notre Dame, USA

of Photoshopped supermodels and celebrities have shown to have a negative effect on the viewer's self-esteem". Some countries have gone so far as to mandate that advertisement photos must declare if the photo is digitally retouched, something commonly termed as the "photoshop law" [2], [3], [4].

As shown in Fig. 1, digital "retouching" can significantly alter facial appearance. If retouched images are used in the biometrics pipeline, recognition accuracy can be considerably affected. In the image forensic domain, this problem is known as "detecting photographic doctoring" [5], [6]. In the biometrics literature, there are two threads which are related to this research: (1) *before-and-after makeup* [7], [8], [9], [16], *disguise* [10] or *plastic surgery*, [11] and (2) *morphing* based synthetic alterations [12].

With readily-available automated tools such as Adobe Photoshop, Corel PhotoPaint, PotraitPro Studio and other packages, it is becoming easy to make sophisticated digitally-retouched versions of original images. Automatic preset parameters for such tools facilitate the process of retouching for users with little or no expertise. Of all the covariates of face recognition, make-up can be considered closest to facial retouching. However, as shown in Fig. 2, the effect of make-up and retouching on facial appearance can be significantly different. For instance, retouching can change the geometric properties of the face by altering the forehead and jaw line or the entire face sculpt. Makeup may make the face look slimmer in some ways, but the face shape in the image remains unchanged. Empirical evidence for digital retouching and makeup having distinct effect on face recognition is also provided in Section V.B in the form of comparison of the proposed retouching detection algorithm with a state-of-the-art makeup detection algorithm [16]. Since retouching changes the appearance, it may also be compared with spoofing. However, we emphasize that retouching should not necessarily be considered as a spoofing [13] or forgery attempt. Generally, there is no malicious intent behind retouching, for instance before uploading to social media websites or magazine covers, images are retouched to improve the appearance. Even though the process deals with creating modified images, it is not done to mis-represent one's identity. However, spoofing or forgery is generally performed when the person is trying to emulate someone else's identity or hide one's own identity.

This research systematically analyzes the effect of retouching on face recognition accuracy. Also, a "deep learning" algorithm is proposed for classifying face images as retouched or original. The contributions of this research are four-fold:

- A large face image database is prepared, the ND-IIITD Retouched Faces database. Original images are taken from the Notre Dame - Set B database [14] and the corresponding retouched images are automatically created using various presets. A celebrity image database (before and after retouching) is also prepared by downloading over 150 image pairs from internet.
- Using the celebrity database, we first establish that digital retouching of face images can seriously degrade the performance of face matchers even when the re-touched (generated from original) image is matched against its original image.
- Next, with the help of controlled experiments on the ND-IIITD Retouched Faces database, we show that different combinations of re-touch operations along with time difference (same session images vs different session images) can have variable effects on performance.
- Finally, we present a novel supervised deep Boltzmann machine learning approach to detecting retouched face images and compare its accuracy with an existing state-of-the-art algorithm [5]. The results demonstrate that the proposed algorithm yields about 39% higher accuracy compared to a previous state-of-the-art [5] algorithm on classifying images as original or retouched.

Fig. 2. Illustrating the difference between makeup and retouching: original images (first column), makeup (second column), and digitally retouched original image (third column). Images in first two columns are obtained from internet; the third image in both rows is retouched by the authors.

TABLE I
DATASET STATISTICS (M- MALE AND F- FEMALE)

Dataset	Subjects	Unaltered	Retouched	Total
Celebrity	165 (M-25, F-140)	165	165	330
ND-IIITD	325 (M-211, F-114)	2600	2275	4875

II. RETOUCHED FACE DATABASES

Since the effect of digitally retouched or altered face images has not been well studied and there is no comprehensive database, we first create two databases and document the baseline performance of face recognition with and without retouching.

A. Celebrity Database

It has been observed that images of celebrities that appear in media collections are generally retouched or modified. The Celebrity database is collected by downloading image pairs from the web. Millions of celebrity images are available online, but it is generally not known whether they are retouched or not. However, some image pairs have been published with asserted ground truth of with and without alteration. We have collected such annotated image pairs for this database. The Celebrity database thus contains two images per person, one that is original and the second that is a retouched version of the original. In total, the database contains 330 images corresponding to 165 individuals. Table I summarizes the statistics of the database and Fig. 3 shows sample images of with and without retouching. Looking closely at these images, it can be observed that after retouching, the skin texture is smoothed, the contrast is enhanced, and facial features are better defined.

B. ND-IIITD Retouched Faces Database

To have a larger, controlled dataset where the retouched/un-retouched ground truth is known, we have used a professional software, PortraitPro Studio Max, to apply various preset retouching operations to known un-retouched (original) face images from the University of Notre Dame database, Collection B [14]. These retouching operations alter facial characteristics such as skin texture, shape of eyes, nose, lips and overall face, prominence of smile, lip shape, and eye color. As shown in Table II, we have created seven different presets representing typical retouching applied to female face images and a different seven presets representing typical retouching for male face images. 325 individuals from the UND-B database that have at least eight images are selected. The first image is kept un-altered to be used as the gallery image for matching experiments. A different retouching is applied to each of the remaining seven images of each subject, using the retouching presets appropriate to the subject's gender. The database contains 2600 original images and 2275 altered images. Table I summarizes the statistics of the ND-IIITD Retouched Faces database and sample male and female images are shown in Figs. 4 and 5 respectively. The figures also show the heat map (visualization) of intensity difference between the original and retouched images. It is evident from the heat maps that major alterations are performed in facial features and the seventh preset contains significant changes in the texture, skin tone and geometry. Instructions to download the ND-IIITD Retouched Faces dataset will be available at http://www.nd.edu/~cvrl.

III. The Effect of Face Retouching on Recognition

To determine the effect of retouched images on face recognition, we have performed four sets of experiments, one on the celebrity database and three on the ND-IIITD Retouched Faces database. The recognition results are computed using Commercial-Off-The-Shelf (COTS) SDK, a commercial matcher that has performed well in NIST evaluations, and OpenBR (v1.1- with default pre-trained engine) [15], which is an open source face recognition engine.

- *Experiment 1*: The original image from the celebrity database is taken as the gallery and the altered image is taken as the probe. Identification experiments are performed and the results are reported in terms of rank-1 identification accuracy.
- *Experiment 2*: One original image from session 1 of the ND-IIITD Retouched Faces database is used as gallery. The remaining seven samples of the ND-IIITD Retouched Faces database are used as probe, once without retouching and once with retouching by applying one preset on each image. This simulates the scenario when the gallery image is captured in an environment such that it is assured to be original, whereas probe images are obtained from sources where images are retouched. Identification accuracies are computed for every preset individually and rank-1 identification accuracies are reported.
- *Experiment 3*: This evaluates the performance when both gallery and probe are retouched images. This simulates the auto-tagging scenarios where both gallery and probe are altered. The gallery images are retouched using a separate preset and then matched with the retouched probe images (with seven presets from Experiment 2).
- *Experiment 4*: This evaluates a scenario where we take two different original (unaltered) images of a person, and get their matching accuracy, and then re-touch both with the same presets, and compare the performance with the original matching results. This is motivated by the idea that a celebrity would retouch all of their images to have the consistent look that they want. On ND-IIITD Retouched Faces database, gallery images are retouched with preset #7 (as shown in Figs. 4 and 5). First, the gallery image is matched with the original (unaltered) 7^{th} image (i.e. two original images are matched) and then the altered gallery image is matched with the altered 7^{th} image (i.e. two altered images with same presets are matched).

The results of the first experiment appear in Table III. The baseline of matching the same image as gallery and probe naturally gives 100% accuracy. However, matching a retouched image as probe to its un-retouched version as gallery results in substantial loss in accuracy. The accuracy of OpenBR drops to 88.73% and COTS matcher drops to 92.45%. *Since the base image used for performing the alterations is same as the gallery image, the drop of 11.27% and 7.55% can be attributed entirely to the presence of retouching*. Fig. 6 shows samples of incorrect rank-1 identification results. The expression, mouth, and nose regions in the best-matched images are more similar compared to the actual gallery image. We have observed that retouching not only focuses on the skin texture but can also alter the geometry to a certain extent, changing the jaw line to make the face somewhat longer or smaller, or decreasing the width of the face to make it thinner. With such variations, it becomes challenging for face recognition algorithms to recognize the images.

Similar analysis is performed on the ND-IIITD Retouched Faces database as well. It is a controlled experiment and the extent of alteration is varied across the seven presets. As shown in Figs. 4 and 5, the first few presets have minor variations whereas, the later presets have significant alterations. It can also be asserted that the alterations which severely affect the geometry and texture information (such as presets 5, 6, and 7) have higher impact on face recognition.

Fig. 3. Sample images from the Celebrity database. The first row contains original images and the second row contains retouched images. Images are from (First pair) http://identity-mag.com/before-after-photoshop/, (Second, Third, and Fourth pairs) http://socioscene.com/here-are-20-celebrities-before-and-after-photoshop-what-the/, and (Fifth pair) http://www.qiwong.com/category/retouching/. Site last accessed on June 22, 2015.

TABLE II
DETAILS OF THE SEVEN MALE AND SEVEN FEMALE PRESETS FROM PORTRAITPRO STUDIO MAX TOOL.

Gender	Preset ID	Characteristics
Female	Preset 1	eyes widen, smooth lips, smile, eyes blue, nose shorten, skin smoothing, skin blush, jaw sculpt
	Preset 2	plump lips, red lips, eyes brighten, eyes brown, nose slim, skin-lighter, skin-blush, hair-enhance color, hair-smooth thicken, forehead-sculpt, face-sculpt
	Preset 3	female standard-full lighting, female-glamourous, red lips, nose-shorten, hair-redden, picture controls-high saturation
	Preset 4	female-skin, lips, eyes only, female-young natural, smile, red lips, eyes brighten, face sculpt, neck lengthen, lighting only-spotlight
	Preset 5	female-skin, lips, eyes only, female-young natural, eyes brighten, eyes widen, nose-slim, red lips, skin-blush, jaw sculpt, neck-lengthen
	Preset 6	plump lips, red lips, eyes widen, eyes brown, nose-slim, skin warmth, enhance hair color, enhance cheekbones, pic controls-high saturation, pic controls-HDR
	Preset 7	smile, plump lips, smooth lips, eyes brighten, eyes green, nose-slim, blush, enhance hair color, smooth thicken hair, hair darken, jaw sculpt, forehead sculpt, face sculpt
Male	Preset 1	male standard-full lighting, plump lips, eyes widen, nose slim,smoothen wrinkles, hair-smooth thicken, forehead-sculpt
	Preset 2	male-glamourous, smile, whiten teeth, smooth lips, eyes brighten, eyes blue, skin smoothing, skin darken, hair redden, neck lengthen
	Preset 3	male young glamorous, plump lips, eyes widen, eyes green, nose shorten, skin smoothing, hair-enhance color
	Preset 4	male-35-45 years glamorous, smooth lips, skin darken, skin warmth, hair-enhance color, hair-smooth thicken, picture control-brighten lighter
	Preset 5	male 35-45 years glamorous, eyes-brighten, plump lips, eyes blue, skin smoothing, hair darken, jaw sculpt, forehead sculpt, enhance cheekbones
	Preset 6	male 50+ years natural, smile, whiten teeth, eyes widen, eyes brown, nose slim, smooth wrinkles, skin lighten, hair redden, face sculpt, neck lengthen
	Preset 7	male 50+ glamorous, male-eyes, lips, skin only, eyes blue, nose-shorten plithrum, hair lighten, jaw sculpt, forehead sculpt, neck lengthen, lighting only-spotlight, rembrandt left

Experiment 2 evaluates the scenario when the gallery image is original whereas the probe images are either (a) original or (b) retouched. On analyzing the results in Table IV, it is observed that the identification accuracy is reduced by retouching for all the presets except preset 1 for COTS matcher, where the accuracy remains same. For the first preset, which only comprises the effect of blush on cheeks and minor skin tone corrections, OpenBR accuracy reduces by only 2.77%. However, for presets 5, 6 and 7 which include thinning the face, changing the brightness and contrast, and other variations, accuracy is reduced by 15.47%, 22.16% and 25.84%, respectively. Note that retouching here includes changes that cannot be achieved with makeup alone.

The third experiment analyzes the performance when both gallery and probe are altered. Here, the probe images are retouched samples from experiment 2 and the gallery images are altered using a new preset, in which some alterations are common to probe presets. The results summarized in Table V show that, for both OpenBR and COTS SDK, the rank-1 accuracies have been affected. It can be inferred that by applying the same presets to all the gallery images, the intra-class variations are increased whereas the inter-class variations

Fig. 4. Sample images from the ND-IIITD Retouched Faces database - male presets. The first row contains original images, the second row contains retouched images, and the third row shows the heat map representation of the intensity difference between corresponding original and retouched images.

Fig. 5. Sample images from the ND-IIITD Retouched Faces database - female presets. The first row contains original images, the second row contains retouched images, and the third row shows the heat map representation of the intensity difference between corresponding original and retouched images.

TABLE III
RANK-1 IDENTIFICATION ACCURACIES ON THE CELEBRITY DATABASE. (THE LAST COLUMN MENTIONS THE DIFFERENCE IN RECOGNITION ACCURACY FROM THE CONSECUTIVE ROWS IN THE SECOND LAST COLUMN)

Algorithm	**Gallery Image**	**Probe Image**	**Accuracy (M-male, F-female)**	**Difference**
OpenBR	Original	Original (Same as gallery)	100 % (M-100%, F-100%)	↓ **11.27%**
		Retouched	88.73% (M-86.36%, F-89.16%)	
COTS	Original	Original (Same as gallery)	100 % (M-100%, F-100%)	↓ **7.55 %**
		Retouched	92.45% (M-88.00%, F-93.28%)	

TABLE IV

RANK-1 IDENTIFICATION ACCURACIES OF FACE RECOGNITION ALGORITHMS ON THE ND-IIITD RETOUCHED FACES DATABASE. THE RESULTS CORRESPOND TO EXPERIMENT 2 WHERE THE GALLERY FOR ALL THE EXPERIMENTS IS THE ORIGINAL UNALTERED IMAGE.

Gallery	Probe	OpenBR accuracy (M- male, F- female)	Change for retouched v. original probe	COTS matcher accuracy (M- male, F- female)	Change for retouched v. original probe
Original Sample	Original Sample 1	96.0 % (M - 100.00%, F - 88.60%)	↓ 2.77%	99.08 % (M - 99.05%, F - 99.12%)	0.00%
	Retouched Sample 1	93.23 % (M - 94.79%, F - 90.35%)		99.08 % (M - 99.05%, F - 99.12%)	
Original Sample	Original Sample 2	93.85 % (M - 92.89%, F - 95.61%)	↓ 7.74%	99.38 % (M - 99.52%, F - 99.12%)	↓ 0.61%
	Retouched Sample 2	86.11 % (M - 82.38%, F - 92.98%)		98.77 % (M - 99.52%, F - 97.37%)	
Original Sample	Original Sample 3	89.54 % (M - 93.36%, F - 82.46%)	↓ 11.39%	99.38 % (M - 99.52%, F - 99.12%)	↓ 0.30%
	Retouched Sample 3	78.15 % (M - 81.04%, F - 72.81%)		99.08 % (M - 99.52%, F - 98.25%)	
Original Sample	Original Sample 4	85.85 % (M - 90.52%, F - 77.19%)	↓ 6.47%	99.38 % (M - 99.05%, F - 100%)	↓ 0.92%
	Retouched Sample 4	79.38 % (M - 72.04%, F - 92.98%)		98.46 % (M - 99.05%, F - 97.37%)	
Original Sample	Original Sample 5	89.85 % (M - 91.00%, F - 87.72%)	↓ 15.47%	99.69 % (M - 99.52%, F - 100%)	↓ 1.54%
	Retouched Sample 5	74.38 % (M - 71.90%, F - 78.95%)		98.15 % (M - 98.10%, F - 98.25%)	
Original Sample	Original Sample 6	83.08 % (M - 83.89%, F - 81.58%)	↓ 22.16%	99.08 % (M - 99.05%, F - 99.12%)	↓ 4.00%
	Retouched Sample 6	60.92 % (M - 61.14%, F - 60.53%)		95.08 % (M - 96.68%, F - 92.11%)	
Original Sample	Original Sample 7	80.92 % (M - 83.41%, F - 76.32%)	↓ 25.84%	98.46 % (M - 98.10%, F - 99.12%)	↓ 3.38%
	Retouched Sample 7	55.08 % (M - 56.87%, F - 51.75%)		95.08 % (M - 92.89%, F - 99.12%)	

TABLE V

RANK-1 IDENTIFICATION ACCURACIES OF FACE RECOGNITION ALGORITHMS ON THE ND-IIITD RETOUCHED FACES DATABASE. THE RESULTS CORRESPOND TO EXPERIMENT 3 WHERE GALLERY IMAGES ARE ORIGINAL AND RETOUCHED AND PROBE IMAGES ARE RETOUCHED.

Gallery	Probe	OpenBR accuracy (M- male, F- Female)	Change for retouched v. original gallery	COTS matcher accuracy (M- male, F- Female)	Change for retouched v. original gallery
Original Sample	Retouched Sample 1	93.23 % (M - 94.79%, F - 90.35%)	↑ 0.62%	99.08 % (M - 99.05%, F - 99.12%)	↓ 0.31%
Retouched Sample		93.85 % (M - 100.00%, F - 82.46%)		98.77 % (M - 99.05%, F - 98.25%)	
Original Sample	Retouched Sample 2	86.11 % (M - 82.38%, F - 92.98%)	↓ 0.31%	98.77 % (M - 99.52%, F - 97.37%)	↓ 1.55%
Retouched Sample		85.80 % (M - 89.52%, F - 78.95%)		97.22 % (M - 99.05%, F - 93.86%)	
Original Sample	Retouched Sample 3	78.15 % (M - 81.04%, F - 72.81%)	↑ 1.23%	99.07 % (M - 99.52%, F - 98.25%)	↓ 1.24%
Retouched Sample		79.38 % (M - 74.89%, F - 87.72%)		97.83 % (M - 97.13%, F - 96.49%)	
Original Sample	Retouched Sample 4	79.38 % (M - 72.04%, F - 92.98%)	↓ 4.61%	98.46 % (M - 99.05%, F - 97.37%)	↓ 2.77%
Retouched Sample		74.77 % (M - 75.36%, F - 73.68%)		95.69 % (M - 99.05%, F - 85.96%)	
Original Sample	Retouched Sample 5	74.38 % (M - 71.90%, F - 78.95%)	↑ 2.46%	98.15 % (M - 98.10%, F - 98.25%)	↓ 2.47%
Retouched Sample		76.84 % (M - 65.71%, F - 97.37%)		95.69 % (M - 96.19%, F - 92.98%)	
Original Sample	Retouched Sample 6	60.92 % (M - 61.14%, F - 60.53%)	↑ 8.31%	95.08 % (M - 96.68%, F - 92.11%)	↓ 6.16%
Retouched Sample		69.23 % (M - 63.51%, F - 79.82%)		88.92 % (M - 92.42%, F - 75.44%)	
Original Sample	Retouched Sample 7	55.08 % (M - 56.87%, F - 51.75%)	↑ 5.84%	95.08 % (M - 92.89%, F - 99.12%)	↓ 4.62%
Retouched Sample		60.92 % (M - 66.82%, F - 50.00%)		90.46 % (M - 83.89%, F - 92.11%)	

Fig. 6. Sample results of incorrect matches from the Celebrity database. The first column shows the probe image (retouched), the second column is the rank-1 match from the gallery (original image of a different individual), and the third column is the correct gallery image (original image corresponding to the input probe).

are reduced. Therefore, the rank-1 accuracies of matching retouched with retouched (when the retouching styles are different) are quite low. Each of these three experiments shows that digital retouching can significantly degrade the accuracy of face recognition algorithms.

Finally, the fourth experiment evaluates the effect when both gallery and probe images are altered using the same style (same preset). As discussed previously, preset #7 is the most challenging for both males and females, and therefore, in this experiment, preset #7 is used for retouching both gallery and probe images. When the two original images are matched, the recognition accuracy from COTS SDK is 98.46% whereas, when the images are retouched with same preset, the accuracy increases to 99.08%. This experimental result implies that if different images are retouched using same style, the intra-class variation is reduced; different images of the same person are made more alike. In this case, the accuracy result can be misleading compared to "real" images.

Fig. 7. Images showing that every celebrity has a distinct and consistent style of their appearance. Images are downloaded from internet.

These experiments justify our initial hypothesis and show that digital retouching can have significant impact on the performance of face recognition algorithms. On the ND-IIITD Retouched Faces database, we applied 14 different presets and the same presets were applied on all the individuals (7 for male and 7 for female). Such a situation can be observed in passport photos, where the photo studios modify or enhance the image in a certain manner that might result in increasing the inter-class similarity. Another aspect of digitally altered photographs is related to celebrities, where everyone has a preference towards a certain kind of appearance. As shown in Fig. 7, every celebrity has their own preferred "look" or "image" that is how they generally want to appear in public or online. This can be ensured either by using makeup or by digital alterations. This kind of tailoring in one's looks can potentially show increased face recognition accuracy. The effects documented in our work suggest that databases of celebrity images from "the wild" naturally incorporate several challenging effects, and that the accuracies obtained on such databases may not necessarily reflect state-of-the-art in unconstrained face recognition on "real" images.

IV. Distinguishing Original and Retouched Images

The results in the previous section suggest that retouched images may not be helpful for important applications such as identity documents. There are applications, such as enforcing the 'photoshop law', where accurate automatic detection of retouched images is important. Therefore, we propose a Supervised Restricted Boltzmann Machine (SRBM) based algorithm to detect whether an image is retouched.

A. Supervised Restricted Boltzmann Machine (SRBM)

In classical form, Restricted Boltzmann Machine (RBM) is unsupervised and follows the energy function

$$E(\mathbf{x},\mathbf{h};\theta) = -\sum_i \sum_j \frac{x_i}{\sigma_i} w_{ij} h_j - \sum_i \frac{(x_i - b_i)^2}{2\sigma_i^2} - \sum_i a_j h_j \tag{1}$$

where $\mathbf{x} \in \mathbf{R}^D$ denotes real-valued input vector (visible layer vector), $\mathbf{h}$ is the hidden layer representation, and $\theta = \{\mathbf{a}, \mathbf{b}, \mathbf{W}, \sigma\}$ are the model parameters. The joint distribution over $\mathbf{x}$ and $\mathbf{h}$, and marginal distribution over $\mathbf{x}$ are

$$P(\mathbf{x},\mathbf{h}) = \frac{1}{Z} exp\left(-E(\mathbf{x},\mathbf{h};\theta)\right) \tag{2}$$

and

$$P(\mathbf{x}) = \sum_h P(\mathbf{x},\mathbf{h}) \tag{3}$$

where, $Z = \sum_{\mathbf{x},\mathbf{h}} exp(-E(\mathbf{x},\mathbf{h}))$ is a partition function. Equation 3 can be written as,

$$P(\mathbf{x}) = \frac{1}{Z} \sum_h exp\left(-E(\mathbf{x},\mathbf{h};\theta)\right) \tag{4}$$

In generative training of RBM, marginal distributions of the visible units $P(\mathbf{x})$ are modeled and the following loss function is utilized [18].

$$L_{gen} = -\sum_{i=1}^{n} logP(\mathbf{x_i}) \tag{5}$$

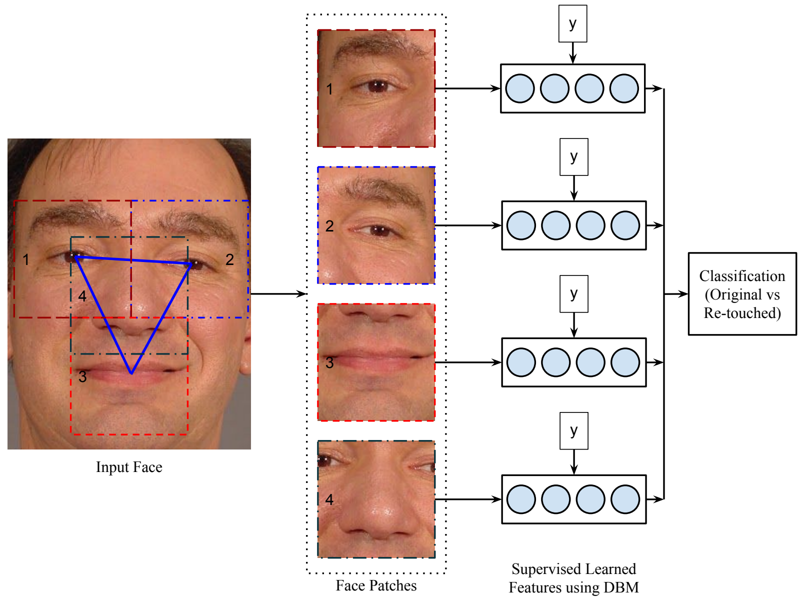

Fig. 8. Steps involved in the proposed retouching detection algorithm.

$\mathbf{X} = \{\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, ...\mathbf{x_n}\}$. Here, n is the number of unlabeled training samples.

Let the labeled training data be $\mathbf{X_c} = \{\mathbf{x_1^1}, \mathbf{x_2^1}, \cdots, \mathbf{x_n^1}, \mathbf{x_1^2}, \mathbf{x_2^2}, \cdots, \mathbf{x_n^2}, \cdots, \mathbf{x_1^m}, \mathbf{x_2^m}, \cdots, \mathbf{x_n^m}\}$ where, m represents the class labels and n represents the number of samples in each class. In order to utilize this labeled training data and make RBM discriminative, one approach is to introduce $l_{2,1}-$norm regularization such that

$$L_{sup} = L_{gen} + \lambda \parallel \mathcal{Z}_c \parallel_{2,1} \tag{6}$$

where, $\mathcal{Z}_c = WX^c$ and λ is the regularization parameter. In maximum likelihood learning, we minimize the loss function,

$$argmin(L_{gen} + \lambda \parallel \mathcal{Z}_c \parallel_{2,1}) \tag{7}$$

The hidden layer representation learnt in this manner encodes class-specific features. The proposed approach is different compared to existing discriminative RBMs [17], [18], where existing approaches either learn class-specific independent RBMs or class labels are introduced in the energy function by adding class labels as a layer in the joint modeling. In the proposed algorithm, by utilizing $l_{2,1}-$norm, we are enforcing class-specific sparsity which helps in extracting features which are discriminative. Similar to the traditional unsupervised approach, the proposed supervised RBM can be stacked to form a deep learning framework (e.g. Deep Supervised Boltzmann Machine). Greedy layer-by-layer training is performed to learn the weights and parameters of the supervised RBM.

B. *Detecting Face Retouching using Supervised RBM*

The proposed supervised RBM can be learned using a labeled training database that consists of the two classes original and retouched. In the proposed framework, four local facial patches are used for detecting retouching. As shown in Fig. 8, the right and left periocular, nose and mouth regions are extracted from a full face (using Viola-Jones face and eye, nose, mouth detector). A three layer $(\mathbf{x}, \mathbf{h_1}, \mathbf{h_2})$ supervised deep Boltzmann machine (SDBM) is learned for each facial patch. In the training data, the patches are labeled as original and retouched, and one SDBM is trained for each region; i.e., one for right periocular, one for left periocular, one for nose, and one for mouth. The output features obtained from the corresponding SDBM are concatenated and given as input to a two-class classifier (in this work, an SVM) for classification.

In this paper, a non-overlapping image set of over 12,000 face images and their corresponding retouched images are used to train the four SDBMs. The size of each facial patch is 64×64 and the size of learned representation for each SDBM is 256. Once the features are learnt and concatenated, 2ν-SVM classifier [19] is trained for two class classification. Since the importance of detecting retouched and un-retouched images can vary depending on context, we have used 2ν-SVM instead of C-SVM so that appropriate constraints and cost can be applied to each class while training the model. Fine tuning of SDBMs and SVM classifier training is performed using the training set of the ND-IIITD Retouched Faces database and classification results are computed on the test set. The next section presents the details of the experiments and showcases the empirical results.

V. Results of Retouching Detection

A. *Comparison with Previous State-of-the-Art Algorithm*

The proposed algorithm is evaluated on the ND-IIITD Faces database and the performance is compared with the current state of the art in detecting retouched images, as represented in the algorithm of Kee and Farid [5]. They used geometric, photometric, and perceptual distortions followed by using support vector regression. On a collection of 468 before/after images obtained from online resources, the results showed that

TABLE VI

CLASSIFICATION ACCURACIES (%) FOR ORIGINAL AND RETOUCHED IMAGES USING THE PROPOSED ALGORITHM AND KEE AND FARID'S ALGORITHM ON THE ND-IIITD AND CELEBRITY DATABASE.

Algorithm	**ND-IIITD database**			**Celebrity database**		
	Overall	**Original**	**Retouched**	**Overall**	**Original**	**Retouched**
Kee and Farid [5]	48.8	32.7	71.9	46.8	27.8	65.8
Proposed (Unsupervised DBM)	81.9	74.3	90.9	88.3	81.0	96.8
Proposed (Supervised DBM)	87.1	81.1	93.9	96.2	93.7	98.7

the metric correlates with perceptual judgements and can be used for classifying photographic and photorealistic images.

The classification accuracies of the proposed and existing [5] algorithms are computed as per the following protocol:

1) In order to train the supervised feature extractor (the proposed algorithm), 12,760 non-overlapping images (both in terms of subjects and images) from the CMU Multi-PIE dataset (frontal faces with neutral and smile expressions for all lighting variations) are used as a separate set. For each of these images, random presets are applied once to create 12,760 retouched images. These 12,760 original + 12,760 retouched = 25,520 images along with their class labels (classes being original and retouched) are used to train SDBM feature extractor.
2) The trained feature extractor with 2ν-SVM classifier is then fine-tuned and trained using $\sim$50% images of both the classes of the ND-IIITD Retouched Faces Database. The remaining non-overlapping (with respect to subjects, i.e. 106 males and 57 females) original and retouched images are used for testing.
3) Similar to the proposed approach, the same train-test partitions from the ND-IIITD Retouched Faces Database are used to train and test Kee and Farid's algorithm. Note that, since the existing algorithm does not require training the feature extractor, only train-test partitions from the ND-IIITD Retouched Faces Database (325 subjects) are used.
4) The same model (trained feature extractor and classifier) is used to evaluate the results on the Celebrity database as well.

As shown in Fig. 9 and Table VI, the proposed supervised deep learning based approach yields a classification accuracy of around 87% and is able to obtain $\sim$81% and $\sim$94% classification accuracies for original (positive) and retouched (negative) classes respectively. The proposed algorithm focuses on four facial patches and supervised features are learnt via deep learning framework to discriminate between original/unaltered and retouched variations. This helps in classifying the test images accurately. On the ND-IIITD database, Kee and Farid's algorithm yields a correct classification accuracy of slightly less than 50% - around 32% original/unaltered images are correctly detected and around 72% retouched/altered images are detected.

Using the trained feature extractor and classifier, when tested on the celebrity database, a classification accuracy of 96.2% is obtained whereas, Kee and Farid's algorithm yields an accuracy of 46.8%. These results show the efficacy of the proposed algorithm in detecting facial retouching and suggest that it can be used in real world applications.

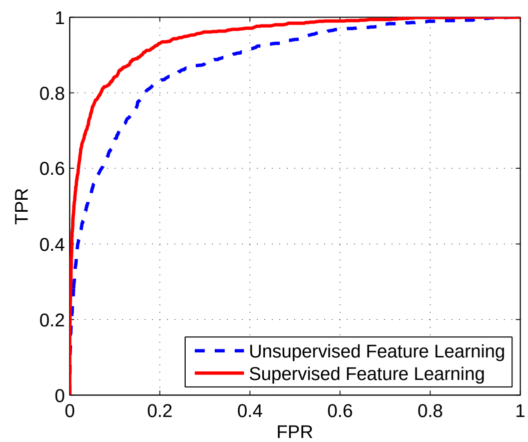

Fig. 9. Results of the proposed supervised DBM based retouching algorithm across different presets.

In order to understand the effect of supervised deep learning over traditional unsupervised feature learning, the performance is compared with traditional DBM. Using the same four patch based approach, in place of SDBM, traditional DBM is used and classification accuracy is computed with the same experimental protocol. Fig. 9 and Table VI show that the supervised version of DBM is about 6% better than the unsupervised counterpart. We have also evaluated the performance of C-SVM by comparing it with 2ν-SVM. On the ND-IIITD database, C-SVM classification yields the overall accuracy of around 84.2% whereas 2ν-SVM yields over 87.1%. We have observed that the main improvement is due to improved classification of unretouched images.

To further understand the classification performance and the effect on the face recognition accuracy, we computed the accuracy for individual retouching presets. On relating these accuracies with the face identification accuracies reported in Experiment II (Table IV), the later presets have more impact on the identification accuracies and therefore, it is more important that these presets are correctly detected. The bar graph in Fig. 10 shows that presets 3 to 7 yield more than 95% classification accuracy whereas there are more errors in the first two presets. As can be observed in Fig. 4 and Fig. 5, the first two presets have very little retouching and also that it is only in local regions. It is our assertion that this can be one of the reasons for higher classification error in the first two presets. As observed in the heat-map (third row of Fig. 4 and Fig. 5) of the remaining five presets, the changes are more prominent and global, and the algorithm yields significantly higher accuracies. Similar performance trend is observed in Kee and Farid's algorithm as well where presets 1 to 4 yield

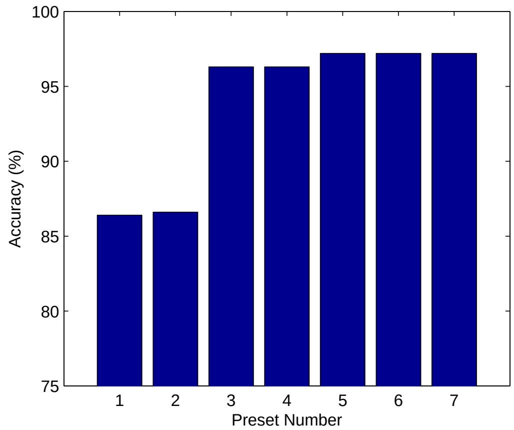

Fig. 10. Results of the proposed supervised DBM based retouching algorithm.

higher errors and presets 5 to 7 are correctly detected. Finally, it is observed that the proposed retouching detection algorithm yields similar classification accuracies on retouched images of males and females, and at 0.05 significance level, the results are statistically not different.

B. Comparison with Makeup Detection

One of the alterations which can be expected to yield a similar effect on faces is *makeup variations*. In order to verify if a state-of-the-art algorithm for makeup detection could solve the problem of retouching detection, the recent algorithm proposed by Kose et. al [16] is evaluated. The algorithm described in the paper uses Local Gabor Binary Patterns (LGBP) and Histogram of Gradient (HOG) for extracting features. The obtained features are normalized and concatenated to obtain the final feature vector. The authors report performance using two classifiers: linear SVM and Alligator, a tool with combination of classifiers. Due to the unavailability of authors' implementation of the algorithm, we have re-implemented their algorithm as described in [16].

Kose et. al [16] have used several databases including YMU (604 images), VMU (204 images), and MIW (154 images) datasets for reporting the performance. Of the existing publicly available makeup datasets, the one that is most relevant to our work is the VMU (Virtual MakeUp) dataset which is prepared using a virtual makeup software. Further, none of these database are large enough to train the proposed retouching detection algorithm. Therefore, in order to perform comparison utilizing same training data, we have trained the existing algorithm [16] on the same training partition of the ND-IIITD Retouched Faces dataset. In this way, the classifiers of the proposed and existing algorithms are trained on the same training set. Testing is performed on the test partition of ND-IIITD Retouched Faces dataset (images from 163 subjects), YMU, MIW, and VMU databases. In this experiment, the existing makeup detection algorithm yields 40 - 55% classification accuracy on these four test databases with the highest accuracy of 55.7% classification accuracy on ND-IIITD Retouched Faces database. Lower performance on makeup databases may be attributed to cross dataset training-testing. On the other hand, the proposed algorithm achieves $\sim$87% classification accuracy on ND-IIITD Retouched Faces database and almost 100% accuracy on the makeup databases. It is our assertion that there is a difference in the problem of makeup detection and retouching detection. Retouching, as presented in this paper, includes changes to the geometry of the face as well as altering the texture. On the contrary, as shown in Fig. 2, makeup variations may make the face look slimmer and texture is "smoothed" but face (or features) shape and dimensions are actually not altered. Further, we also believe that the existing makeup detection algorithms may not address the problem of retouching detection as efficiently as the proposed deep learning architecture based algorithm.

VI. CONCLUSIONS

Users upload millions of face images every day onto social media sites. Before uploading, they often digitally retouch their photos. In digital marketing, retouched face images are likely the *rule* and un-retouched (i.e. original) face photos are an *exception*. If users supply their own images for identification documents, these may be retouched as well. In experiments with two databases prepared by the authors, retouching clearly is shown to degrade the accuracy of face recognition. When both the enrolled image and the probe image of a person are retouched versions of different original images, the estimated accuracy can be higher than the accuracy for matching the un-retouched images. This result has implications for face recognition studies that are based on image datasets of celebrities collected from the web. When the enrolled image is un-retouched and the probe image is retouched, or vice-versa, the matching accuracy can be greatly reduced in comparison to the situation when both images are un-retouched. Thus the problem of classifying images as retouched/un-retouched is suggested as a challenging but important problem.

We propose a novel supervised deep learning based algorithm to solve the problem of classifying face images as original or retouched. The proposed algorithm shows a significant improvement compared to state-of-the-art algorithm for retouching detection. Additional experiments show that the improvement in classification accuracy can be attributed to the supervised DBM and to the form of the SVM used for classification. Experimental results are also reported on several different face makeup databases, suggesting that the makeup detection problem is substantially different and may be less challenging than the problem of detecting digitally retouched face images.

ACKNOWLEDGEMENT

Aparna Bharati would like to thank Saket Anand and Angshul Mazumdar for their valuable feedback on this work.

REFERENCES

[1] IBT News, *Supermodels Without Photoshop: Israel's Photoshop Law Puts Focus On Digitally Altered Images*, http://goo.gl/3PvaEf, Last Accessed May 9, 2015.

[2] Reuters News, *France bans super-skinny models in anorexia clampdown*, http://www.reuters.com/article/2015/04/04/us-france-anorexia-idUSKBN0MU0JK20150404, Last Accessed May 9, 2015.

[3] US Congress *Truth in Advertising Act of 2014*, https://www.congress.gov/bill/113th-congress/house-bill/4341, Last Accessed May 9, 2015.

[4] Digiday News, *There's a push to make photoshopped models in ads illegal*, http://digiday.com/brands/truth-in-advertising/, Last Accessed May 9, 2015.

[5] E. Kee, and H. Farid, *A perceptual metric for photo retouching*, P.N.A.S., vol. 108, pp. 19907–19912, 2011.

[6] E. Kee, J. F. Obŕien, and H. Farid, *Exposing Photo Manipulation from Shading and Shadows*, ACM Transactions on Graphics, vol. 33, pp. 165:1–165:21, 2014.

[7] A. Dantcheva, C. Chen, and A. Ross, *Can facial cosmetics affect the matching accuracy of face recognition systems?*, IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems, 2012, pp. 391–398.

[8] M. Ferrara, A. Franco, D. Maltoni and Y. Sun, *On the Impact of Alterations on Face Photo Recognition Accuracy*, International Conference on Image Analysis and Processing, 2013, pp. 743–751.

[9] C. Chen, A. Dantcheva and A. Ross, Automatic facial makeup detection with application in face recognition, International Conference on Biometrics, 2013, pp. 1–8.

[10] T. Dhamecha, R. Singh, M. Vatsa and A. Kumar, *Recognizing Disguised Faces: Human and Machine Evaluation*, PLoS ONE, vol. 9, pp. e99212, 2014.

[11] R. Singh, M. Vatsa, H. S. Bharadwaj, A. Noore and S. S. Nooreyezdan, *Plastic Surgery: A New Dimension to Face Recognition*, IEEE Transactions on Information Forensics and Security, vol. 5, pp. 441–448, 2010.

[12] M. Ferrara, A. Franco and D. Maltoni, *The Magic Passport*, IEEE/IAPR Proceedings of International Joint Conference on Biometrics, 2014, pp. 1–7.

[13] S. Marcel, M. S. Nixon, and S. Z. Li. Handbook of Biometric Anti-Spoofing. Springer, 2014.

[14] P. J. Flynn, K. W. Bowyer, and P. J. Phillips, *Assessment of time dependency in face recognition: An initial study*, Audio and Video-Based Biometric Person Authentication, 2003, pp. 44–51.

[15] J. C. Klontz, B. F. Klare, S. Klum, A. K. Jain, and M. J. Burge, *Open source biometric recognition*, IEEE International Conference on Biometrics: Theory, Applications and Systems, 2013, pp. 1–8.

[16] N. Kose, L. Aprville and J.L. Dugelay, *Facial Makeup Detection Technique Based on Texture and Shape Analysis*, 11th IEEE International Conference on Automatic Face and Gesture Recognition, 2015, pp. 1-7.

[17] H. Larochelle and Y. Bengio, *Classification using discriminative restricted Boltzmann machines*, 25th International Conference on Machine learning, 2008, pp. 536-545.

[18] T. Schmah, G. E. Hinton, S. L. Small, S. Strother, and R. S. Zemel. *Generative versus discriminative training of RBMs for classification of fMRI images*, In Advances in neural information processing systems, pp. 1409-1416, 2008.

[19] H. G. Chew, C. C. Lim and R. E. Bogner, *An implementation of training dual-nu support vector machines*, Optimization and Control with Applications, Qi, Teo, and Yang, (eds.), 2004, Kluwer.

Aparna Bharati received her undergraduate degree from the Indraprastha Institute of Information Technology (IIIT) Delhi, India in 2015. Currently, she is a doctoral student at the University of Notre Dame. Her research interests include image processing, pattern recognition, machine learning, and biometrics.

Richa Singh (S'04 - M'09 - SM'14) received the M.S. and Ph.D. degrees in computer science from West Virginia University, Morgantown, USA, in 2005 and 2008, respectively. Her research has been funded by UIDAI and DeitY, Government of India. She is currently an Associate Professor with the Indraprastha Institute of Information Technology, Delhi, India. She has authored over 150 publications in refereed journals, book chapters, and conferences. Her areas of interest are biometrics, pattern recognition, and machine learning. She is a member of the Committee to Define an Extended Fingerprint Feature Set, the Computer Society, and the Association for Computing Machinery. She is a recipient of the Kusum and Mohandas Pai Faculty Research Fellowship at the Indraprastha Institute of Information Technology, the FAST Award by DST, India, and several best paper and best poster awards in international conferences. She is also an Editorial Board Member of Information Fusion (Elsevier) and the EURASIP Journal on Image and Video Processing (Springer). She is serving as the General Co-Chair of ISBA2017 and PC Co-Chair of BTAS 2016.

Mayank Vatsa (S'04 - M'09 - SM'14) received the M.S. and Ph.D. degrees in computer science from West Virginia University, Morgantown, USA, in 2005 and 2008, respectively. His research has been funded by UIDAI and DeitY, Government of India. He is currently an Associate Professor and AR Krishnaswamy Faculty Research Fellow with the Indraprastha Institute of Information Technology, Delhi, India. He has authored over 150 publications in refereed journals, book chapters, and conferences. His areas of interest are biometrics, image processing, computer vision, and information fusion. He is a member of the Computer Society and the Association for Computing Machinery. He is a recipient of the FAST Award by DST, India, and several best paper and best poster awards in international conferences. He is also the Vice President (Publications) of IEEE Biomertics Council, an Associate Editor of the IEEE ACCESS and an Area Editor of Information Fusion (Elsevier). He served as the PC Co-Chair of ICB 2013, IJCB 2014, and ISBA2017.

Kevin W. Bowyer is the Schubmehl-Prein Professor of Computer Science and Engineering at the University of Notre Dame and also serves as Chair of the Department. Professor Bowyer's research interests range broadly over computer vision and pattern recognition, including data mining, classifier ensembles and biometrics. Professor Bowyer received a 2014 Technical Achievement Award from the IEEE Computer Society, with the citation "For pioneering contributions to the science and engineering of biometrics'.

Over the last decade, Professor Bowyer has made numerous advances in multiple areas of biometrics, including iris recognition, face recognition, and multi-biometric methods. His research group has been active in support of a variety of government-sponsored biometrics research programs, including the Human ID Gait Challenge, the Face Recognition Grand challenge, the Iris Challenge Evaluation, the Face Recognition Vendor Test 2006, and the Multiple Biometric Grand Challenge. Professor Bowyer's most recent book is the Handbook of Iris Recognition, edited with Dr. Mark Burge.

Professor Bowyer is a Fellow of the IEEE, a Fellow of the IAPR, and a Golden Core Member of the IEEE Computer Society. Professor Bowyer is serving as General Chair of the 2015 IEEE International Conference on Automatic Face and Gesture Recognition. He has previously served as General Chair of the 2011 IEEE International Joint Conference on Biometrics, as Program Chair of the 2011 IEEE International Conference on Automatic Face and Gesture Recognition, and as General Chair of the IEEE International Conference on Biometrics Theory Applications and Systems in 2007, 2008, 2009, and 2015. Professor Bowyer has also served as Editor-in-Chief of the IEEE Transactions on Pattern Analysis and Machine Intelligence and Editor-In-Chief of the IEEE Biometrics Compendium, and is currently serving on the editorial board of IEEE Access.