Texture Analysis-based Makeup Removal for Imposture Identification in Individual Biometric Validation

R.Logeswari Saranya [1*], K Umamaheswari [2]

1 Assistant Professor(SS), Department of Artificial Intelligence and Data Science, Dr NGP Institute of Technology, Coimbatore, India.
[e-mail:logeswarisaranya@drngpit.ac.in]

2 Professor, Department of Information Technology, PSG College of Technology, Coimbatore, India.
[e-mail.com:umakpg@gmail.com]

*Corresponding author:logeswarisaranya@drngpit.ac.in

Abstract - Face recognition is rapidly becoming one of the most popular biometric authentication methods. Most face recognition systems are focused on extracting features and enhancing their verification and identification capabilities. The detection of security vulnerabilities of different types of attacks has been given attention only in recent years. These attacks can include, but are not limited to: Obfuscation Spoofing and morphing for example, a hacker can masquerade as a target to gain access to the biometric system. The application of cosmetics can alter the appearance of a face leading to a decreased characteristic distinctiveness of faces. Facial makeup includes variations in skin tones, the position of eyebrows and skin complexion. The cosmetic effect on an individual causes the face recognition system to falsely identify the person affecting the security of the biometric system. Adding a presentation attack detection module to the existing biometric system can be the solution to this problem. In this work, a CNN-based machine learning approach is adapted to classify the presentation attack using texture analysis. The proposed method is to extract the original face by removing makeup so that the FR system recognizes the person's real identity, resulting in decreased vulnerability. The false accept rate (FAR) is a measure of a biometric system’s resistance to zero-effort attacks and is generally considered as the system’s performance.

Keywords- Makeup Detection, Makeup Removal, Convolutional Neural Network, GAN models, Haar-cascade Algorithm

1. Introduction

The biometric system is the security system that recognizes and identifies people based on theirbiological and behavioral characteristics. Thisbiometric authentication system is used for various applications like the security of computers and mobile phones, airports, banks, military bases, biometric attendance, and tracking systems. Though biometric systems improve security, like any other system,they are vulnerable and prone to threats.

Face recognition is one of the popular biometric authentication methods. The system is said to be vulnerable when it falsely identifies an individual and gives them access rights which can lead to the exploitation of information. This vulnerability is due to a variety of attacks like spoofing, obfuscation, morphing, and makeup. The solution to the spoofing attacks has already been studied in lots of papers.

Presentation Attack Detection (PAD), also referred to as “Anti-spoofing” or “Liveness detection,” may be a critical capability to think about when deploying face recognition in automated authentication and identification scenarios. Whereas face recognition determines if a presented face matches a registered record, PAD determines whetherthe face itself may be authentic or is a copy of the face, from a photograph to a video sample on an LCDto a high-resolution 3D mask.

Despite a good deal of progress in face recognition systems, vulnerabilities to face spoof attacks are mainly overlooked. The facial spoof attack may be a process during which a fraudulent user can subvert or attack a face recognition system by masquerading as a registered user and thereby gaining illegitimate access and advantages. Face spoofing attacks may be a major issue for companies selling face biometric-based identity management solutions. It is thus essential to develop robust, efficient, and compact face anti-spoofing (or liveness detection) methods, which are capable of generalizing well to discriminative, class-specific information and imaging conditions.

To achieve this goal, the main focus is on the attack caused by makeup, which is the makeup attack. The application of cosmetics on the face to look like other people like celebrities can sometimes confuse the biometric system. Makeup can cause variation in skin tone, skin complexion, lip color, eye shadows, the position of eyebrows, and the overall appearance of the person. This vulnerability due to the makeup attack must be addressed to improve the biometric system security. The overall flow of paper goes as first have gone through all the survey papers and got the limitations on the existing system. Second proposed structure is defined. Third experimental result and evaluation is done. Finally comparison is done between models.

2. Literature Survey

Several papers and applications on makeup detection and facial makeup removal have been studied as a reference for this work. The details that have been inferred from those applications and papers are discussed below.

2.1 Makeup Classification

The proposed method uses a feed-forward back-propagation neural network-based classifier for classification. The classification makes use of features extracted using a discrete wavelet transform approach from face samples [4]. The most commonly used neural network architecture with the back propagation algorithm is the multilayer feed-forward network. This technique is not only computationally less extensive as compared with other techniques but also provides the best result on various images. The robustness for image variation in rotations, illuminations, etc must be improved. The evaluation of the robustness of thelargest data sets is necessary for practical use.

The vulnerability of a widely used open-source face recognition system (i.e.) Arc face, to makeup presentation attacks using makeup-induced face spoofing datasets like MIFS and FRGCv2.The success rate of makeup attacks in the MIFS datasets has an impact on the security of the face recognition system. The warping technique is used to simulate improved makeup presentation attacks which provide a higher success rate [7]. The m-PAD technique is used to compare and classify the bonafide and makeup images. Provides better performance in all the datasets. A convolutional neural network is used to distinguish between presentation with age-induced facial makeup and without makeup. Proposed presentation attack detection provides a 6.6% Average classification error rate in the AIM (Age-induced makeup) dataset and 4% in all the datasets. AIM dataset contains 200+ video presentations of old-age makeup and original faces each [3]. AIM dataset results in a 14% decrease in the median matching scores of recent CNN-based FR systems. Overall accuracy is 93% using the AIM dataset.

2.2 Makeup Removal

The WGAN-GP approach is used to remove the makeup. The dataset collected consists of five separate datasets (MIFS, FAM, YMU, VMU, MIW) of a total of 2600 images of 1300 different people. Each person has two images one with makeup and another one without makeup. CNN model is developed to classify whether the person is with makeup or without makeup and then built a generative adversarial network (GAN) model to remove the makeup from the image [6]. The best accuracy obtained for this model was 80% on training and 79% on testing.

The aim is to promote the existing verification system to accept or reject the claimed identity of a person with makeup in an image. A makeup robust face verification framework is proposed based upon a generative adversarial network. The proposal synthesizes non-makeup face images from makeup images. Specifically, a patchwise contrastive loss is introduced in the generative model to constrict the distance between makeup and non-makeup images [5].

A bidirectional tunable de-makeup network (BTD-Net) is proposed for removing makeup effects. For tractable learning of the makeup process, which is one-to-many mapping determined by the cosmetics that are applied, a latent variable is used which reflects the style of the makeup [1]. This latent variable is extracted from the de-makeup process and used as a condition of the makeup process to constrain the one-to-many mapping to a specific result. The proposed BTD-Net surpassed the state-of-the-art techniques in estimating realistic no-makeup faces that correspond to the input makeup images.

3. Proposed Method for Imposture Identification

The proposed system in which two major level of processing is done first classification and the second is removal. In classification phase first image take as an input may consists of noise which should be removed. Gabor filter is used to remove noise. Second pre processing is done in order to extract the features of the image. Finally based on feature the different convolution neural network hidden layers are used to classify the categories of makeup and no makeup images. In removal phase generative adversarial network is used to generate real image from fake images in a cyclic fashion.

3.1 Work Flow of Makeup Classification

3.1.1 Data Pre-processing

Data Pre-processing is one of the important steps in machine learning. Each image in the dataset (both makeup and no makeup images) is processed before being given as input to the model. The face from the image is detected using the Haar-cascade algorithm. The detected face is then cropped and saved as an image. By doing this, the images that contain no face are eliminated and only the face is extracted leaving the unwanted information as shown in Figure 7.At first, the input images are resized into 256x256 and the bilateral filter is applied to remove any available noise. To extract the features, the Gabor filter is applied to the image. This process is shown in Figure 8.The image is then divided into 9*9=81 blocks. Uniform LBP is calculated for each of these blocks. A histogram is computed for each block and is then concatenated to form a feature vector as shown in Figure 1.

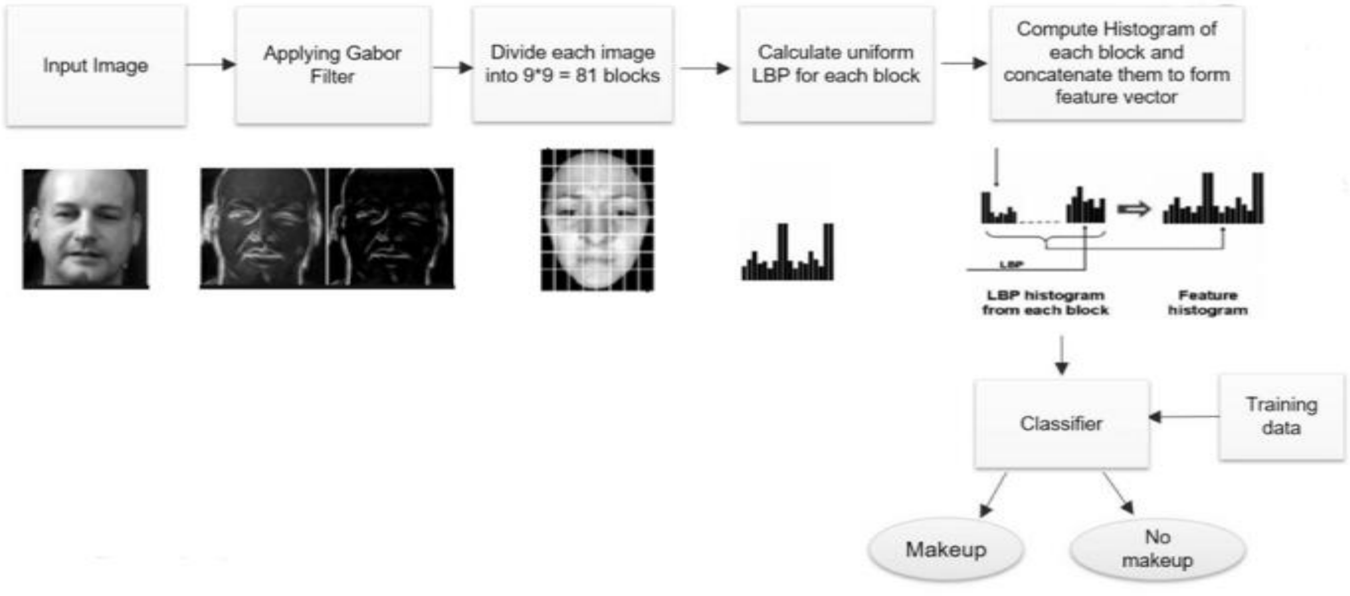

Input Image
Applying Gabor Filter
Divide each image into 9*9 = 81 blocks
Calculate uniform LBP for each block
Compute Histogram of each block and concatenate them to form feature vector
LBP
LBP histogram from each block
Feature histogram
Classifier
Training data
Makeup
No makeup

Figure 1. Workflow of Classification Model

3.1.2 Step by Step process of proposed VGG Architecture

The proposed deep convolution neural network VGG consists of several layers to classify the makeup and on makeup image. The steps as follows

1. First the network is characterized by its simplicity, using only 3 × 3 convolutional layers piled on top of each other to add depth.

2. Second 48 x 48 grayscale image is given as input to the first convolutional layer.

3. Third the image is passed through a stack of five convolutional layers each with a kernel size of 3x3. Each convolution layer results in the output of feature maps of that image which is fed as input to the next layer.

4. Every convolutional layer is followed by the rectified linear activation function (ReLU).

5. The down-sampling of feature maps is done using a max-pooling layer over a 2x2 pixel window.

6. The flattening layer is used to convert 2-dimensional feature maps into a single long feature vector, whichis connected to the final classification model.

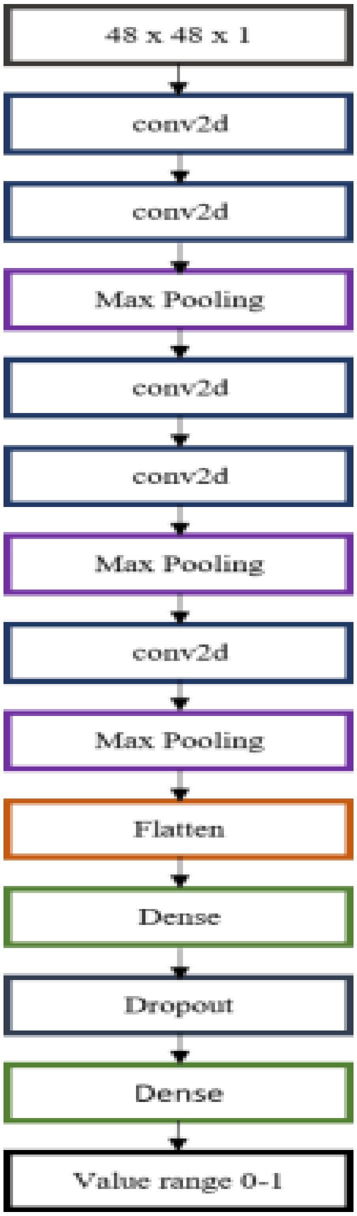


Figure 2. Proposed VGG Architecture

The extracted feature is fed to the classifier model which classifies the given input image as makeup or no makeup one. The classifier is built using the VGG architecture mentioned in Figure 2. This classifier returns a value that ranges from 0 to 1. Since the sigmoid activation layer is used, the value below 0.75 indicates that the image belongs to class 0 which is no makeup and the value above 0.75 indicates thatthe image belongs to class 1 which is makeup.

3.2 Work Flow for Makeup Removal Phase

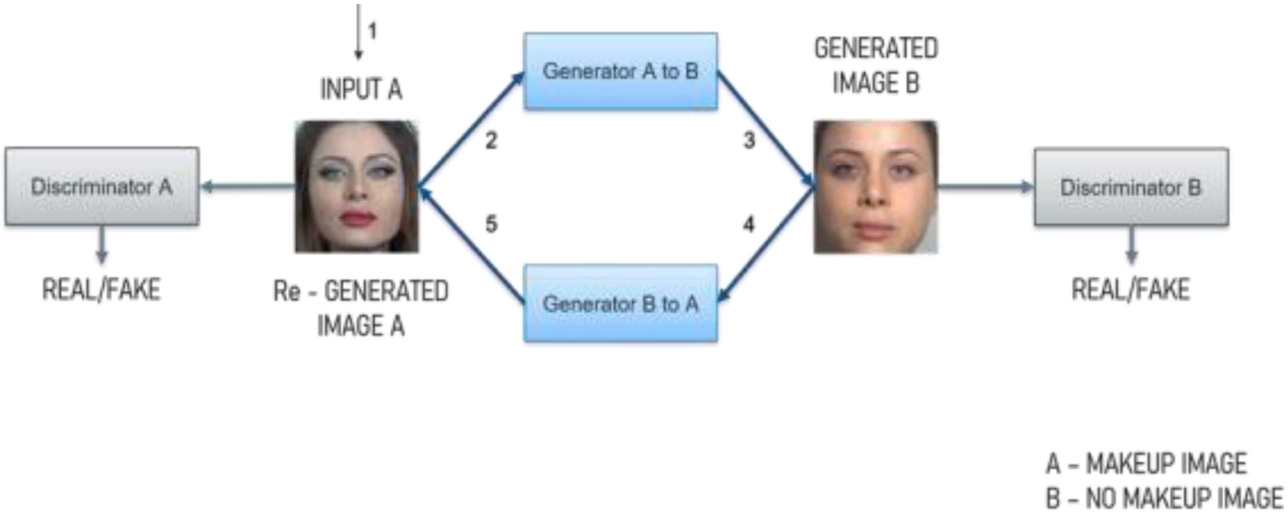


Figure 3. Workflow of Makeup Removal Model

The proposed makeup removal model uses generative adversarial network which works like a cyclic process. Back tracking can be done easilier using the reverse mechanism. The working of generators and discriminators as follows

- Generator A takes makeup images as input and generates the no-makeup images.
- Generator B takes the no-makeup images from generator A as input and generates makeup images.
- The generated images are then passed to the discriminator models which check the plausibility of the images and update the generator models accordingly as shown in Figure 3.

Here only the makeup to no makeup translation is required. Hence, we concentrate on generator A. The makeup image is given as an input to this model which generates the corresponding no-makeup image. After detecting whether the subject is wearing makeup or not using the presentation attack classification model, this makeup removal model is used to extract the original face. This can prevent the vulnerability caused by makeup attacks and improve the security of biometric systems.

3.2.1 Step by Step process of proposed pix2pix GAN

The proposed Pix2pix GAN is based on the conditional generative adversarial network, where a target image is generated. The condition is placed on the given input image.

- First the discriminator is provided both with a source image and the target image and must determine whether the target is a plausible transformation of the source image.
- Second generator is trained via adversarial loss, which encourages the generator to generate plausible images in the target domain.
- The generator is also updated via L1 loss measured between the generated image and the expected output image.
- Finally additional loss encourages the generator model to create plausible translations of the source image.

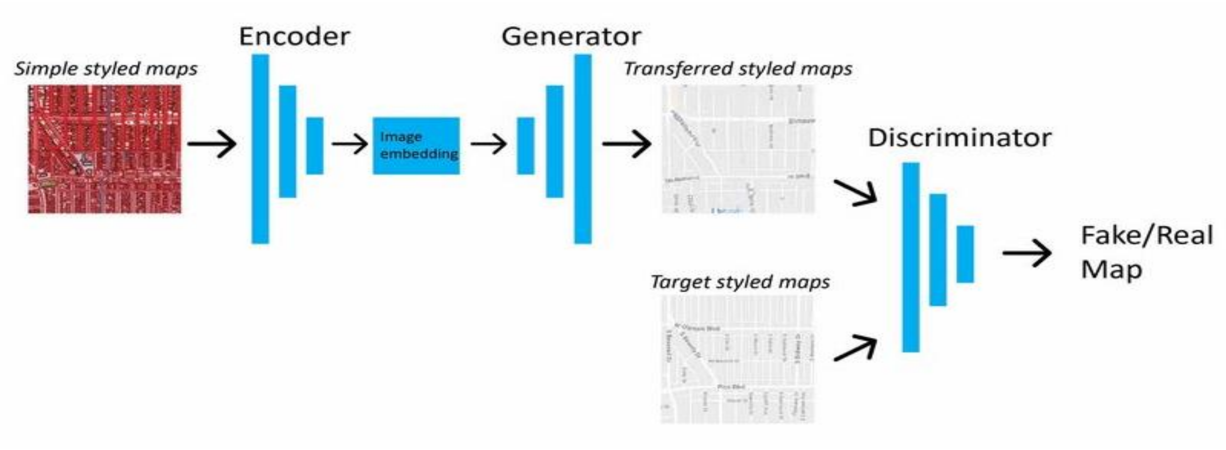


Figure 4. Encoder and Decoder of pix2pix GAN

The encoder and decoder of the generator are comprised of standardized blocks of convolutional, batch normalization, dropout, and activation layers as shown in Figure 4. This standardization means that we can develop helper functions to create each block of layers and call it repeatedly to build up the encoder and decoder parts of the model. Major limitation of Pix2pix GAN is, it accepts only paired images as input for training the model. This problem is solved using CycleGAN.

3.2.2 Step by Step process of proposed CycleGAN

This Cycle GAN is a model is for focusing unpaired images. In this proposed Cycle GAN involves in training of two generator models and two discriminator models. The process is given as follows

- The discriminator model is implemented as the PatchGAN model. PatchGAN is mainly used for image translation.
- The discriminator model takes 256x256 RGB images as input. Discriminator uses theblocks of conv2D, Instance normalization, and leakyRelu layers.
- A sequential model was created using CNN layers.
  - In the first layers, it takes 4X4 convolution layers with 64 filters and 2x2 strides, and instance normalization is not used in the first layer instead it takes 0.2 as slope.
  - In the Second layer, it takes 4x4 convolution layers with 128 filters and 2x2 strides, and in the third layer, it takes 4x4 convolutionlayers with 256 filters and 2x2 strides.
  - Finally, we compile this model for 4 epochs using the Adamoptimizer.
- The generator model takes 256x256 RGBimages.
  - In the first layer, it takes 7x7 convolution layers with 64 filters and 2x2 strides, and Instance Normalization is used in the first layer.
  - In the second layer, it takes 3x3 convolution layers with 128 filters and 2x2 strides.
  - In the third layer, it takes 3x3 convolutional layers with 256 filters and 2x2 strides. Residual blocks are used in generators which are mainly for image transformation

3.3 Loss Functions to Improve Performance

To improve the performance of the model four loss functions have been implemented. These loss functions are explained below.

3.3.1 Adversarial loss

The adversarial loss is calculated based on the probabilities returned by the discriminator network. In the adversarial model, the discriminator network is fed with generated images generated by the generated network.

$$l^{SR}_{Gen} = \sum_{n=1}^{N} -logD_{\theta_D}(G_{\theta_G}(I^{LR}))$$

Here, $G_{\theta G}(I^{LR})$ is the generated image and $D_{\theta D}(G_{\theta G}(I^{LR}))$ represents the probability that the generated image is a real image.

- Adversarial loss - In Adversarial loss, the makeup image is given as an input to Generator B which generates the no makeup image. The no makeup image is given to discriminator B which should discriminate the image as real or fake one.

3.3.2 Identity Loss

Loss of identity is added to maintain the tone. It says that if the generator receives an image of the target class, it should return the same image.

F(x) ≈ x and G(y) ≈ y.

λ is a term added to define the relative importance of cycle and identity losses, compared to the GAN losses

- Identity loss - In Identity loss, the no-makeup image is given as an image to Generator B which should give the same no-makeup image as the output.

3.3.3 Cycle Loss

From left to right: input, cycle consistency loss alone, adversarial loss alone, GAN + forward cycle loss (F(G(x)) ≈ x, labels→photos) GAN + backward cycle loss (G(F(y)) ≈ y, photos→labels, CycleGAN (ours), and ground truth. Both Cycle alone and GAN + backward fail to produce images similar to the target domain.

- Forward cycle loss - In forward cycle loss, the makeup image is given as an input to Generator B, which generates the no makeup image. When this no-makeup image is given to Generator A, it should generate the same makeup image. The Forward cycle loss follows the order 2->3->4->5 as shown in Figure 3.
- Backward cycle loss - The backward cycle loss is the reverse process of the forward cycle loss. It follows the order 4->5->2->3as shown in Figure 3.

3.4 Data Collection

For imposture identification, the dataset should contain makeup and no makeup images. Due to limited dataset available, dataset such as MIFS (Makeup Induced Face Spoofing), YMU (YouTube Makeup dataset), VMU (Video Makeup dataset), FAM (FAce Makeup), MIW (Makeup In Wild) are combined together. This combined dataset is split into two categories - (i) Makeup and (ii) No makeup images. The dataset contains 949 makeup images and 1085 no makeup contains images as shown in Figure 5 and Figure 6.

3.5 Steps in Training the model

1. The batch size is fixed at one image. Since the dataset has 949 makeup images, the batches per epoch will be 949 i.e., for one epoch 949 training iterations will be done.
2. A batch of real images and fake images from both domains (makeup and no makeup) is generated and the fake images are updated to the discriminator's fake image pool.
3. Then for each iteration, both the generator and discriminator will be trained over one batch of samples, and the model will be saved.

3.6 Steps in Testing the model

1. The model from makeup to no makeup is loaded.

2. The input image is resized and normalized and it will be passed to the loaded model.
3. The model will generate the no-makeup image and it will be plotted.

4. Experimental Result

First Session

4.1 Sample Images from Dataset

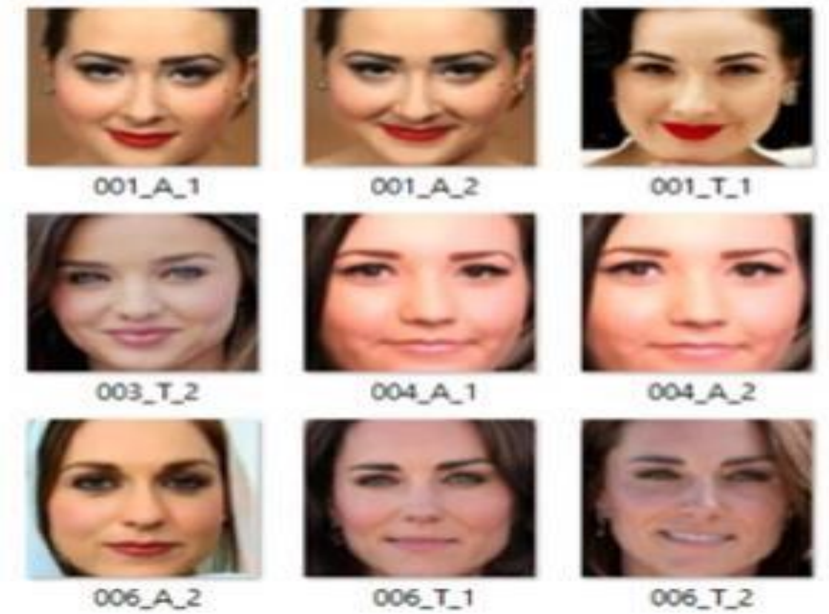


Figure 5. Sample Makeup Images

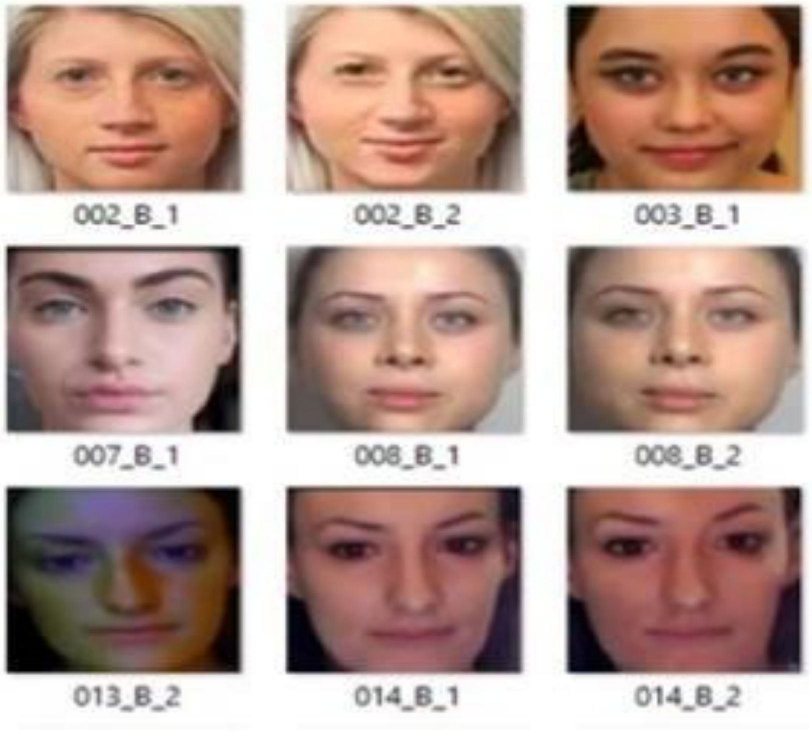


Figure 6. Sample No Makeup Images

4.2 Data Pre-processing

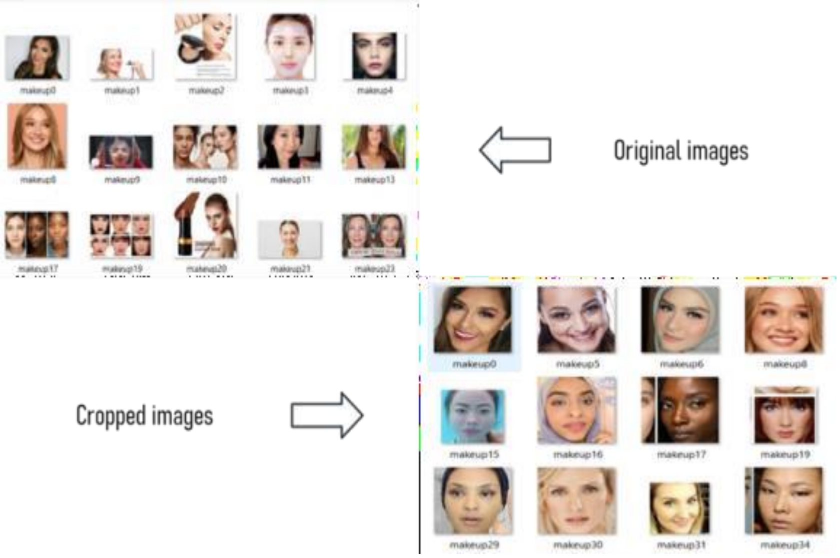


Figure 7. Pre-processed Cropped Images from Original Images

Images are converted into pixel values and stored in an array. Then normalization is applied to speed up the convergence.

4.3 Feature Extraction in Makeup Classification Model

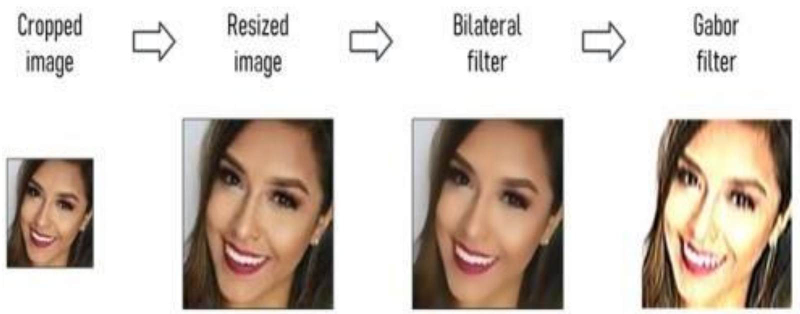
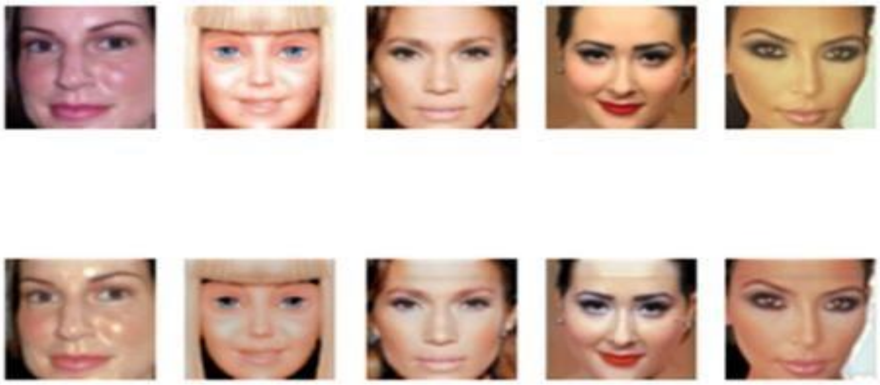


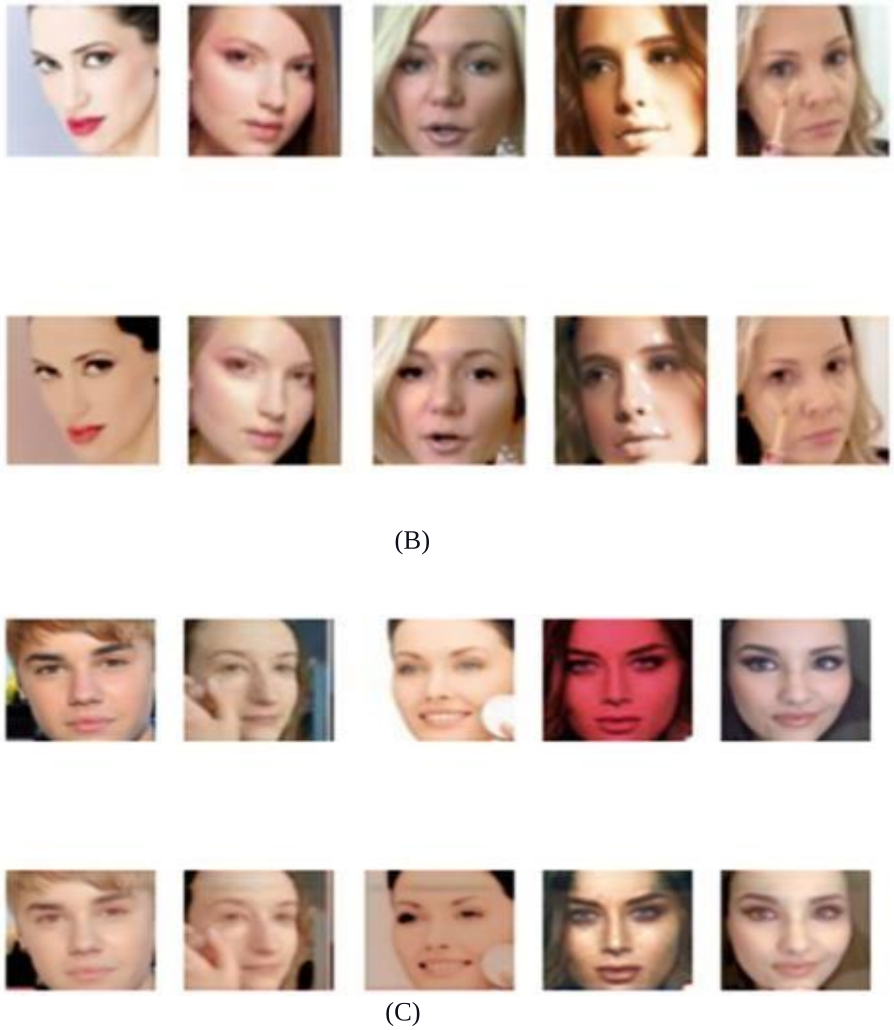
Figure 8. Feature Extraction using Gabor Filter

4.4 Sample Result for Makeup Removed Images from Dataset

The detected makeup image from the classification model is passed into the makeup removal model to get the no makeup image as the output.



(A)



(B)

(C)

Figure 9. Makeup to No Makeup Generation (of each image A, B, and c the first row contains makeup images and the corresponding image after makeup removal is in the second row)

4.5 Analysis of Evaluation Metric in Makeup Classification Model

The classification model is used to identify whether the person is wearing makeup or not. When an image is passed to this model it outputs a value ranging from 0 to 1.The performance of the classification model is evaluated in terms of accuracy, loss, precision, recall, and f1 score

Table 4.1 Evaluation of Makeup Classification Model

Evaluation Metrics	**Training Size=90%** **Testing Size=10%**
Accuracy (%)	97.48
Precision (%)	97.44
Recall (%)	96.59
F1 Score (%)	95.39

4.6 Comparison of Proposed GAN Models

By training both pix2pix GAN and CycleGAN and comparing their image quality measure [2] them as shown in Table 4.2., it is proved that cycleGAN is more efficient than pix2pix GAN.

Table 4.2 Comparing the Quality Measure of GAN Models

GAN Models	SSIM (Structural Similarity Index Measure)	FSIM (Feature Similarity Index Measure)	FID (Frechet Inception Distance)
Pix2pix	42.86	69.34	19.50
Cycle	48.70	75.48	39.48

5. Conclusion

Thus by implementing this method to the existing biometric system, the real identity of the person can be identified though the person intends to fraudulently access the system by wearing makeup. Thus, the vulnerability caused by makeup attacks can be prevented which improves the security of the biometric system. The captured image from a live videocamera is passed to the makeup classification model to identify whether the person is wearing makeup or not. After this, the identified makeup image is passed to the makeup removal model to extract the bare face.

To improve this model further, instead of an image dataset, video samples can be collected. It will be more appropriate for the biometric system since it captures real-time instances. The security of the biometric system can be increased by preventing not only the makeup attack but also other presentation attacks like spoofing and morphing which are more likely. So, a combination of all the attack prevention models can be developed.

References

[1] Nat´alia Machado Anchieta, Anthonieta Looman Mafra, Roberta Tokumori Hokama, Marco Antonio Correa Varella, Jailson de Almeida Melo, Luana Oliveira da Silva, Caio Santos Alves da Silva, and Jaroslava Varella Valentova "Makeup and Its Application Simulation Affect Women's Self-Perceptions. Archives of Sexual Behavior", 50(8):3777–3784, 2021, https://doi.org/10.1007/s10508-021-02127-0.

[2] Herng Hua Chang and Chun Hsiao Yeh "Face anti-spoofing detection based on multi-scale image quality assessment", Image and Vision Computing, 121:104428, 2022, https://doi.org/10.1016/j.imavis.2022.104428.

[3] Yuting Du, Tong Qiao, Ming Xu, and Ning Zheng, "Towards Face Presentation Attack Detection Based on Residual", Color Texture Representation. Security and Communication Networks, 2021, https://doi.org/10.1155/2021/6652727.

[4] J V Gorabal and D H Manjaiah, "Texture Analysis for Face Recognition", International Journal of Graphics and Multimedia (IJGM), 4(2):20–30, 2013.

[5] Dhiman Karmakar, Puja Mukherjee, and Madhura Data, "Spoofed Facial Presentation Attack Detection by Mul-tivariate Gradient Descriptor in Micro-Expression Region", Pattern Recognition and Image Analysis, 31(2):285–294,2021, doi: 10.1134/S1054661821020097.

[6] C. Rathgeb, P. Drozdowski, and C. Busch, "Detection of makeup presentation attacks based on deep face representations", Proceedings - International Conference on Pattern Recognition, pages 3443–3450, 2020, DOI: 10.1109/ICPR48806.2021.9413347.

[7] C. Rathgeb, P. Drozdowski, D. Fischer, and C. Busch, "Vulnerability Assessment and Detection of Makeup Presentation Attacks", 8th International Workshop on Biometrics and Forensics, IWBF 2020 - Proceedings, 2020, DOI: 10.1109/IWBF49977.2020.9107961.