AI

MDPI

Article

Deep Learning Models for Automatic Makeup Detection

Theiab Alzahrani [1], Baidaa Al-Bander [2] and Waleed Al-Nuaimy [1,*]

[1] Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool L69 3GJ, UK; pstalzah@liverpool.ac.uk
[2] Department of Computer Engineering, University of Diyala, Baqubah 32010, Iraq; baidaa.q@gmail.com
* Correspondence: wax@liverpool.ac.uk; Tel.: +44-151-794-4512

Abstract: Makeup can disguise facial features, which results in degradation in the performance of many facial-related analysis systems, including face recognition, facial landmark characterisation, aesthetic quantification and automated age estimation methods. Thus, facial makeup is likely to directly affect several real-life applications such as cosmetology and virtual cosmetics recommendation systems, security and access control, and social interaction. In this work, we conduct a comparative study and design automated facial makeup detection systems leveraging multiple learning schemes from a single unconstrained photograph. We have investigated and studied the efficacy of deep learning models for makeup detection incorporating the use of transfer learning strategy with semi-supervised learning using labelled and unlabelled data. First, during the supervised learning, the VGG16 convolution neural network, pre-trained on a large dataset, is fine-tuned on makeup labelled data. Secondly, two unsupervised learning methods, which are self-learning and convolutional auto-encoder, are trained on unlabelled data and then incorporated with supervised learning during semi-supervised learning. Comprehensive experiments and comparative analysis have been conducted on 2479 labelled images and 446 unlabelled images collected from six challenging makeup datasets. The obtained results reveal that the convolutional auto-encoder merged with supervised learning gives the best makeup detection performance achieving an accuracy of 88.33% and area under ROC curve of 95.15%. The promising results obtained from conducted experiments reveal and reflect the efficiency of combining different learning strategies by harnessing labelled and unlabelled data. It would also be advantageous to the beauty industry to develop such computational intelligence methods.

Keywords: makeup detection; deep learning; convolution neural networks; semi-supervised learning; auto-encoder; self-learning

Citation: Alzahrani, T.; Al-Bander, B.; Al-Nuaimy, W. Deep Learning Models for Automatic Makeup Detection. *AI* **2021**, *2*, 497–511. https://doi.org/10.3390/ai2040031

Received: 9 July 2021
Accepted: 11 October 2021
Published: 14 October 2021

1. Introduction

Facial makeup has a long history. It is an instance of a cosmetic modification that may modify the face's perceived appearance. Ageing, natural biological change, and plastic surgery, a medically induced change, are other forms of alterations. In general, surgical modifications are expensive and permanent. Non-permanent cosmetic improvements, such as makeup, on the other hand, tend to be quick, cost-effective and socially acceptable; at the same time, they can alter appearance considerably. Makeup alterations are aimed at (a) altering the perceived facial form by emphasising contouring techniques; (b) altering the perceived nose shape and size by contouring tools; (c) enhancing or decreasing the perceived mouth size; (d) altering the appearance and contrast of the mouth by adding colour; (e) altering the perceived eyebrow shape, colour and location; (f) altering the perceived form, size and contrast of the eyes; (g) hiding dark circles under the eyes; and (h) altering the texture and colour of the perceived skin. In addition to the effects described above, cosmetics can also be used to effectively disguise and conceal wrinkles, birth moles, scars and tattoos [1].

A developing cosmetics industry aims to enhance and improve facial attractiveness while projecting good health. Makeup or cosmetics encompasses a wide range of methods, categories, products and have become socially acceptable in all aspects of our lives. The use of makeup, on the other hand, presents a significant obstacle to biometric systems. Facial makeup has the ability to change and conceal one's natural appearance, making certain identification and verification functions more difficult [2,3]. An experimental study conducted by [4] revealed that the application of facial makeup leads to changes in the skin texture, smoothness, and colour tone. Many other studies showed that facial makeup carries an inherent ambiguity due to artificial colours, shading, contouring, and varying skin tones. The issue becomes more complicated as makeup changes the symmetry of certain facial features like eyes and lips, affecting the distinctive character of faces [5–11].

Facial recognition technology is already ubiquitous worldwide, where it is used for everything from making payments to catching jaywalkers. Avoiding the cameras might be impossible, but it is possible to prevent them from recognising individuals by manipulating facial features. The manipulation could result from applying makeup that contrasts with natural skin tone anywhere on the region where the nose, eyes, and forehead intersect and using it to modify the dark and light areas of the face. Multi-coloured hairpieces can also act as a powerful tool for disrupting symmetry, especially if they conceal the top half of the face [12–14]. To mitigate the effects of facial disguising, researchers in Facebook have recently developed a de-identification system that can supposedly alter key facial features in videos, even working in real-time for live streams, to decorrelate the identity [15].

With advancements in machine learning and computer vision techniques, deep neural network models are becoming very successful and widespread. Considering that deep learning architectures have been successfully used in various fields, including facial image analysis [16–18], it could even further be exploited to detect the faces disguised by makeup to overcome the flaws in many facial-related analysis methods. Those models have the ability to extract the features directly from images without the need for human interaction by training on large datasets using various types of learning schemes. A trained human expert or a physical experiment are required to label the data for a learning problem. Therefore, the costs associated with the labelling process will make providing a massive and fully labelled training sets a difficult task, whereas collection of unlabelled data is relatively inexpensive. In machine learning, semi-supervised learning defines a type of algorithms that attempt to learn from both unlabelled and labelled examples. Many semi-supervised learning with deep neural networks were designed based on generative models such as denoising auto-encoders [19], stacked convolutional auto-encoders [20], variational auto-encoders [21] and generative adversarial networks [22]. Pseudo-labelling of self-learning is another powerful method for semi-supervised learning that has shown success in the context of deep learning models [23]. Semi-supervised learning with self-learning works by assigning approximate classes on unlabelled data by making predictions from a model trained only on labelled data.

The remainder of this paper is presented as follows: in Section 2, the related work is presented; proposed methods are described and explained in Section 3; the materials and results of the proposed systems are revealed and discussed in Section 4; finally, the work has been concluded in Section 5.

2. Related Work

The beautification effects caused by cosmetics have been studied in the research literature. In comparison to work on makeup recommendations, research involving an image of an already made-up face appears to be extremely rarer. The first study that objectively identified the effect of facial makeup on a face recognition system was performed by Dantcheva et al. [24]. They used three techniques related to face recognition which are Gabor wavelets, local binary pattern, and the commercial Verilook Face Toolkit, to test the accuracy of recognition before and after makeup using two databases: the YouTube MakeUp (YMU) database and the Virtual MakeUp (VMU) database. The presence of

makeup in facial images is also identified in [25] using a feature vector that includes shape, texture, and colour information. Another study, [4], tackled the facial verification issue by extracting features from both a makeup-wearing and a makeup-free face, then matching the two faces using correlation mapping. In [26], an approach for makeup face verification was presented by measuring correlations between face images in a meta subspace where canonical correlation analysis (CCA) and support vector machine (SVM) were used for learning and verification, respectively.

The interrelationship between, on the one hand, the subjective interpretation of female facial beauty and, on the other hand, selected objective parameters including facial features, photo-quality and non-permanent facial features were examined and analysed by the authors in [27]. A supervised deep Boltzmann machine was also proposed by the authors [28] to solve the problem of classifying face images as original or retouched. The authors in [29] used Gabor filtering and histogram of oriented gradients (HOG) methods for feature extraction from VMU and YMU datasets. These features were combined to form the final feature vectors, which were then reduced using the fisher linear discriminant analysis and classified. In [5], the authors presented a locality constrained low-rank dictionary learning algorithm to determine and locate the usage of cosmetics.

Fu and Wang [30] have recently created a system capable of identifying, analysing and digitally removing makeup from a facial image. For makeup detection using a supervised deep learning approach, authors in [31] developed a system based on a pre-trained Alex network for makeup detection. Yi Li et al. [11] proposed learning from generation approach for makeup-invariant face verification by introducing a bi-level adversarial network (BLAN). The vulnerability of many face recognition systems has been recently assessed to facial makeup presentation in [32]. Furthermore, the same research team presented a facial retouch detection method based on analysis of photo response non-uniformity (PRNU) in [33] and conducted a review and benchmarking for makeup presentation attacks methods in [34].

The scenario of makeup detection can be leveraged in many real-life applications. For instance, the presence of makeup can significantly affect the results of beauty and attractiveness prediction systems [3,27,28,30]. The development of makeup detection methods could grant advantages to beauty prediction systems to refine their outcomes by utilising prior knowledge about the makeup presence in a given facial image. Furthermore, detecting the makeup-wearing can benefit automated age estimation systems [35,36] and facial recognition models from the perspective of security [24,25,32,33].

Motivated by the above-reported observations, we introduce makeup detection schemes that help detect the facial images covered by makeup using labelled and unlabelled data. The contribution of our work is two-fold: (i) conducting a comparative study by investigating the impact of different learning strategies to automatically detect the presence of makeup, and (ii) carrying out a thorough analysis and comprehensive study involving six challenging datasets. This enables for evaluation with facial images collected from various sources, demonstrating the robustness and relatability of the methods.

3. Materials and Method

3.1. Materials

Six publicly available datasets are used in this study including Kaggle [37], YMU [24], VMU [25], MIW [38], MIFS [39], FAM [26]. The Kaggle dataset contains facial images of subjects obtained from the Internet comprising 1506 images of people wearing makeup and not wearing makeup. YouTube Makeup (YMU) is a set of 151 facial images collected from YouTube makeup tutorials, focusing on Caucasian females. The images were taken before and after the subjects had their makeup applied. There are four shots per subject: two before makeup application and two after makeup application, totalling 604 images. Virtual Makeup (VMU) is a set of facial images of Caucasian female subjects from the FRGC repository [40]. To simulate the application of makeup, this dataset has been synthetically modified. This change was made using the publicly available software Taaz [41]. The

total number of images in the dataset is 204, with one before-makeup shot and three after-makeup shots per subject. Makeup in the Wild (MIW) is unconstrained face images of 125 subjects with and without makeup obtained from the Internet. There are (1–2) images per subject: one before-makeup and one after makeup producing 154 images (77 with makeup and 77 without makeup). Makeup Induced Face Spoofing (MIFS) is a dataset of 624 images collected from 107 subjects (4 images) and 107 target subjects (2 images). Images of a subject before makeup, images of the same subject after makeup with the intention of spoofing, and images of the target subject who is being spoofed are three image sets included in this dataset. Finally, the FAce Makeup (FAM) dataset contains 519 subjects. Each subject has two face images, one is with makeup, and the other is not, producing 1038 images. All subjects involved in the study gave their consent through a data collection process led by the authors, who generated the data and made it publicly accessible for scientific research purposes. Figure 1 shows some image samples used in this study.

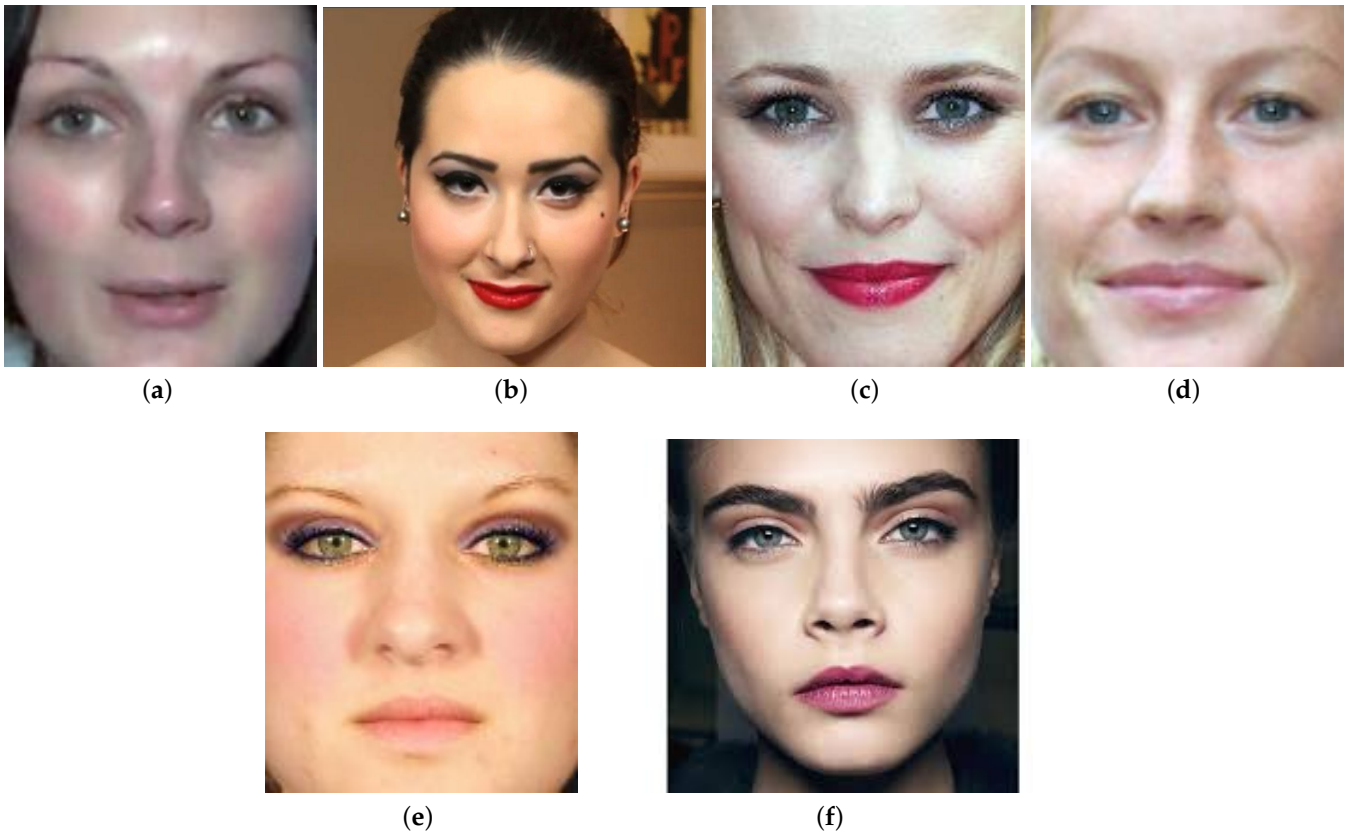


Figure 1. Samples of images from datasets used in the study. (**a**) YMU; (**b**) MIFS; (**c**) MIW; (**d**) FAM; (**e**) VMU; (**f**) Kaggle.

3.2. Method

This work proposes and designs an automated facial makeup detection system leveraging three types of learning schemes. The deep learning models are evaluated and tested on multiple public datasets, and their performances are compared to ground truth annotations. The classifier in the models deals with makeup identification in an image as a two-class detection task; with makeup and without makeup classes. Using labelled and unlabelled data, we harness semi-supervised strategies and transfer learning schemes to implement the proposed systems based on convolutional neural networks (CNNs) as follows:

1. Supervised Transfer Learning of Pre-trained VGG16 CNN: pre-trained VGG16 network [42] is fine-tuned on labelled data to extract the facial features and produce a makeup classifier (absence or presence of makeup).

2. Semi-Supervised CNN with Self-Learning: the fine-tuned VGG16 network resulting from the previous stage can be combined with a self-learning algorithm developed in [23] which is trained on unlabelled data to produce a semi-supervised learning scheme.
3. Semi-supervised CNN with Convolutional Auto-encoder (CAE): in this model, CAE [20] is used to extract the salient visual features in an unsupervised learning manner using the unlabelled makeup data. The trained CAE is then used for initialising the weights of supervised CNN. Whereas the weights of fully connected layers are trained using the labelled data.

3.2.1. Supervised Learning with Pre-Trained CNN

Transfer learning aims to use knowledge from the source task to increase learning in the target task. There are three indicators in which the transfer of knowledge might help boost learning. The first is the initial performance achieved in the target task utilising only transferred knowledge before any further learning. The second factor is the time it takes to learn the target task thoroughly using transferred knowledge versus learning it from scratch. The third factor is the difference between the final performance level achieved in the target task after applying transfer learning and fine-tuning and the final level without transfer. Negative transfer occurs when a method of transfer degrades performance. Positive transfer between adequately related tasks while avoiding negative transfer between tasks that are less correlated is one of the most complex issues in designing transfer strategies. When transfer learning occurs from one task to another, transferring the properties of one task onto the attributes of the other is often required to establish correspondences [43].

Transfer learning strategies harnessed to boost the evaluation performance in many computer vision tasks can be categorised into three categories,

1. Re-training the entire pre-trained CNN architecture on the target dataset, yet avoiding the random weight initialisation by starting from the pre-trained weight values.
2. Training the new classifier (fully connected layers) related to the target task and freezing the other layers (convolutional and all other layers). In this transfer learning strategy, the pre-trained CNN works as a feature extractor where the weights of the CNN layers, except fully connected layers, are retained without change.
3. Training some of the convolutional layers, especially the top layers of CNN, and the classifier (fully connected layers). The original weights are exploited as a starting point for learning.

To conduct knowledge transfer in our comparative study, we leveraged pre-trained convolutional neural network VGG16 [42]. We re-trained the pre-trained VGG16, which is earlier trained on a large-scale hierarchical image database (Imagenet) [44], using our datasets. Figure 2 depicts the transfer learning using pre-trained VGG16 model for makeup detection.

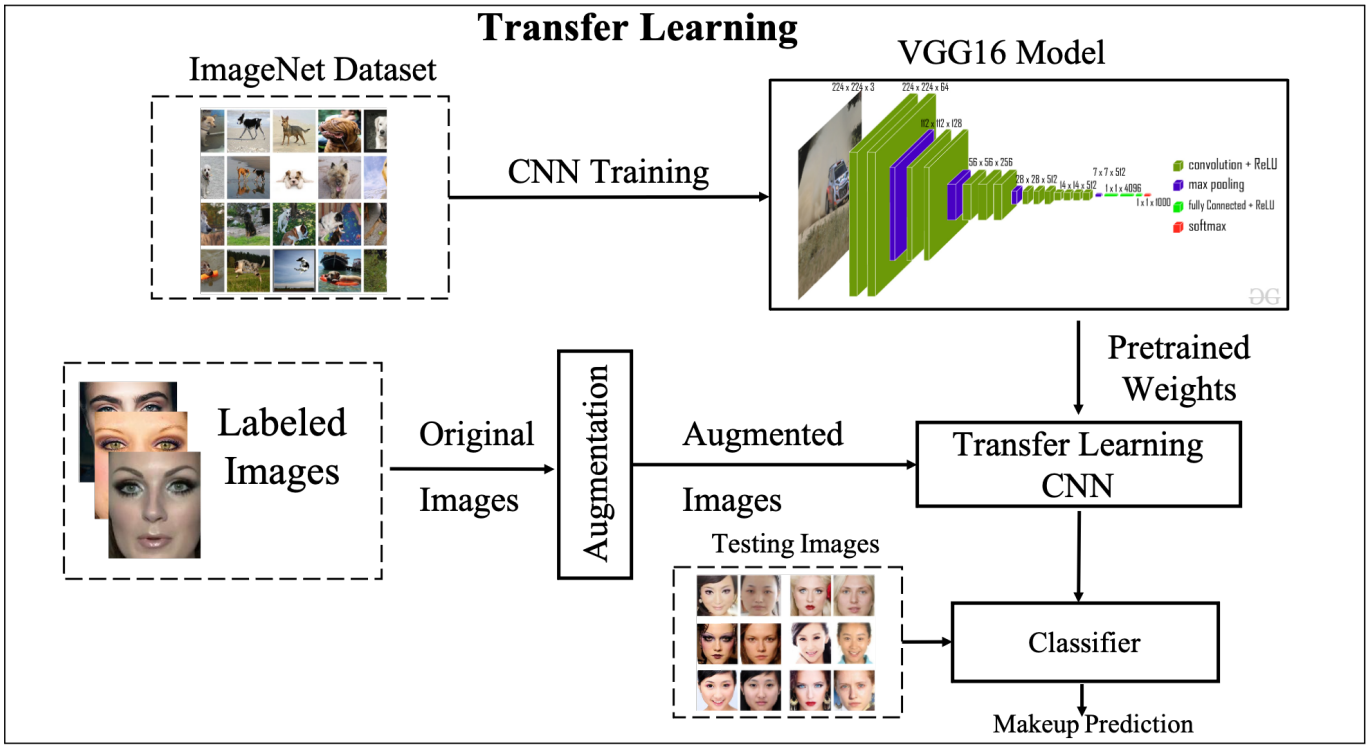


Figure 2. Supervised learning scheme with a transfer learning using pre-trained VGG16 model for makeup detection.

3.2.2. Semi-Supervised Learning with Self-Learning Pseudo-Labelling

Unlike the supervised learning scheme that requires completely labelled data, semi-supervised learning combines supervised and unsupervised learning techniques. As the name indicates, semi-supervised learning uses both a labelled set of training data and an unlabelled collection of training data. The self-training scheme introduced in [23] works by training the classification models (CNNs) on labelled instances and employing the trained classifier to identify class labels of the unlabelled examples. The predicted class labels with the highest probability of being correct are adopted as pseudo-labels. The examples with pseudo-labels are then used to train the classifier. Once the classifier is trained, it can be used to test unseen instances. The description of pseudo-labelling with self-learning merged with classifier is presented in Figure 3.

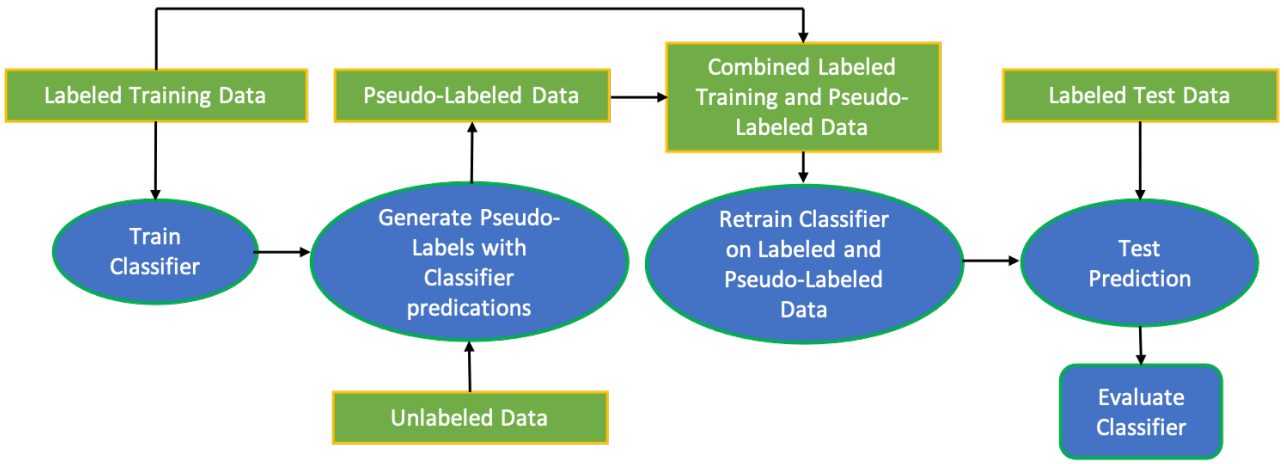


Figure 3. Self-learning with pseudo-labels scheme.

On a conceptual level, this scheme integrates the pre-trained VGG16 model with the self-learning method to double the size of the training dataset by choosing the most reliable examples from the unlabelled image data. The steps of learning process which are repeated until the convergence using makeup labelled images l and an unlabelled images u could be described as follows:

1. The pre-trained VGG16 is used to predict the labels of the unlabelled images u.
2. The prediction scores with highest confidence rates ($l^{'}$ pseudo-labels) obtained from applying VGG16 model are selected and combined with label images l.
3. The combined pseudo-label images and labelled images are then used to train the classifier. The loss function can be represented as $Loss_{total} = Loss_{labelled} + Loss_{unlabelled}$.

The unlabelled image samples u that are left unclassified after the algorithm's convergence are omitted. Figure 4 illustrates the semi-supervised learning scheme with pseudo-labels and self-learning for makeup detection.

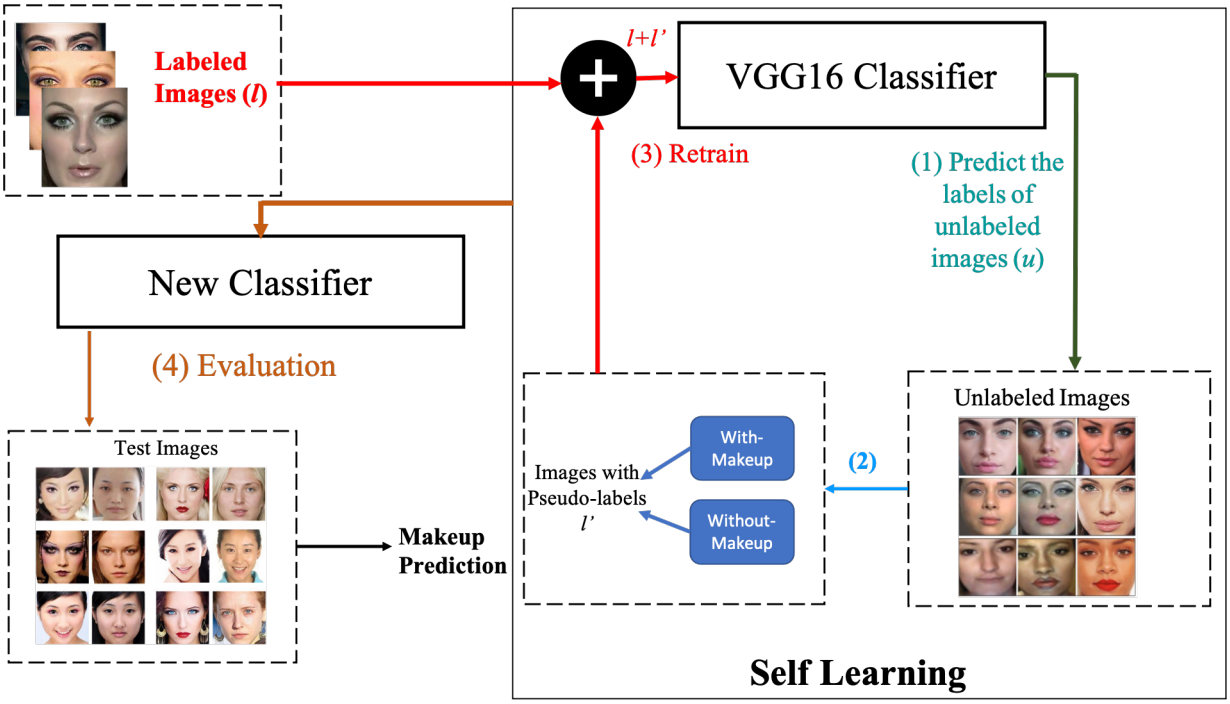


Figure 4. Semi-supervised learning scheme with pseudo-labels and self-learning for makeup detection.

3.2.3. Semi-Supervised Learning with Convolutional Auto-Encoder

Autoencoders [45] are neural networks widely used to extract features (feature learning) and dimensionality reduction. Several auto-encoders can be stacked to form a deep hierarchy constituting a convolutional auto-encoder (CAE). In a greedy, layer-wise manner, unsupervised training can be achieved in CAE. Back-propagation can then be used to fine-tune the weights. The top-level activations can be used as feature vectors for classifiers [20]. Our implemented convolutional auto-encoder (CAE) contains the layers described in Table 1. Our convolutional auto-encoder (CAE) contains an encoder path, latent features, decoder path. The encoder part comprises eight convolution layers, each followed by batch normalisation, max-pooling, and drop out layers. The decoder part consists of eight up convolution layers, each followed by batch normalisation, upsampling, and drop out layers.

The basic mathematical principle of standard fully connected auto-encoder (AE) requires mapping the unlabelled input image x into a latent representation h in the hidden layers using the following formula:

$$h = s(Wx + b) \tag{1}$$

where s is a non-linear activation function, W is the weights of the network, and b is bias. The auto-encoder then tries to decode and reconstruct the unlabelled input image $\hat{x}$ by:

$$\hat{x} = f(W^{'}h + b^{'}) \tag{2}$$

The auto-encoder is learned using a back-propagation algorithm to minimise the reconstruction error. The extended version of the conventional fully connected auto-encoder (AE) is convolutional auto-encoders (CAE) which is a combination between fully connected auto-encoder and convolutional neural networks (CNNs). To integrate CNNs with auto-encoders, a corresponding deconvolutional layer must be built for each convolutional layer. Furthermore, because max-pooling layers cause information loss, an un-pooling layer should resemble the original values. Deconvolution layers are equivalent to the convolutional layers but transposed. Unlike AE, the weights in CAE are shared among all locations in the input image, leading to preserve spatial locality. The latent representation of the k-th hidden layer is represented by:

$$h^k = s(x * W^k + b^k) \tag{3}$$

The reconstructed image is defined by the formula:

$$\hat{x} = f(h * W^{'} + b^{'}) \tag{4}$$

$W^{'}$ identifies the transpose operation of the encoder layer weights, and $*$ refer to the convolution operation. The objective function targeted to be minimised is defined as MSE cost function to measure the error between the original and constructed images. Figure 5 explains the semi-supervised learning scheme with a convolutional auto-encoder for makeup detection.

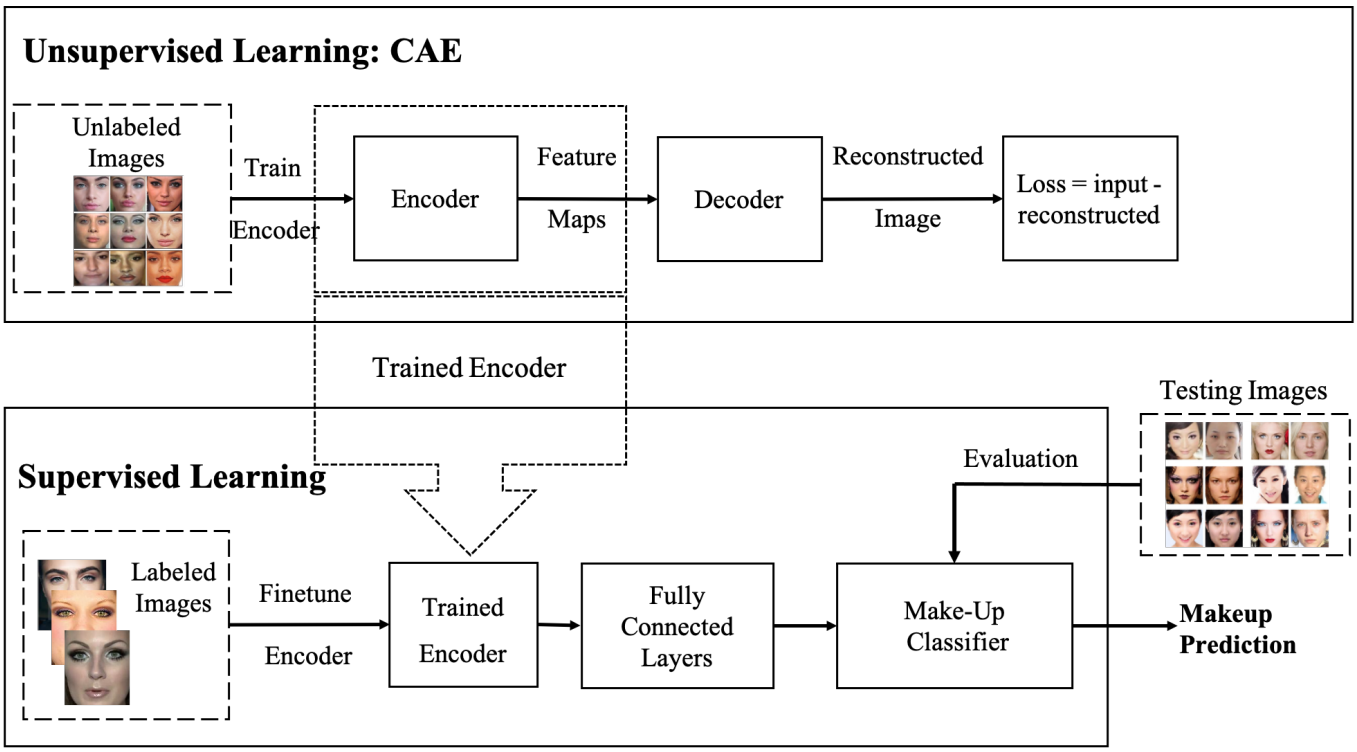


Figure 5. Semi-supervised learning scheme with a convolutional auto-encoder for makeup detection.

Table 1. The architecture of the implemented convolutional auto-encoder.

Layer (Type)	Output Shape	#Param
input-1(InputLayer)	(128, 128, 3)	0
conv-1(Conv2D)	(128, 128, 32)	896
batchnormalisation	(128, 128, 32)	128
conv-1-2 (Conv2D)	(128, 128, 32)	9248
batchnormalisation-1	(128, 128, 32)	128
pool-1(MaxPooling2D)	(64, 64, 32)	0
dropout(Dropout)	(64, 64, 32)	0
conv-2(Conv2D)	(64, 64, 64)	18,496
batchnormalisation-2	(64, 64, 64)	256
conv-2-2(Conv2D)	(64, 64, 64)	36,928
batchnormalisation-3	(64, 64, 64)	256
pool-2(MaxPooling2D)	(32, 32, 64)	0
dropout1(Dropout)	(32, 32, 64)	0
conv-3(Conv2D)	(32, 32, 128)	73,856
batchnormalisation-4	(32, 32, 128)	512
conv-3-2(Conv2D)	(32, 32, 128)	147,584
batchnormalisation-5	(32, 32, 128)	512
pool-3(MaxPooling2D)	(16, 16, 128)	0
dropout-2(Dropout)	(16, 16, 128)	0
conv-4(Conv2D)	(16, 16, 256)	295,168
batchnormalisation-6	(16, 16, 256)	1024
conv-4-2(Conv2D)	(16, 16, 256)	590,080
batchnormalisation-7	(16, 16, 256)	1024
pool-4(MaxPooling2D)	(8, 8, 256)	0
dropout-3(Dropout)	(8, 8, 256)	0
flatten(Flatten)	(, 16,384)	0
latent-feats(Dense)	(, 1024)	16,778,240
reshape(Reshape)	(2, 2, 256)	0
upsample-4(UpSampling2D)	(4, 4, 256)	0
upconv-4(Conv2D)	(4, 4, 256)	590,080
batchnormalisation-8	(4, 4, 256)	1024
upconv-4-2(Conv2D)	(4, 4, 256)	590,080
batchnormalisation-9	(4, 4, 256)	1024
upsample-3(UpSampling2D)	(8, 8, 256)	0
dropout-4(Dropout)	(8, 8, 256)	0
upconv-3(Conv2D)	(8, 8, 128)	295,040
batchnormalisation-10	(8, 8, 128)	512
upconv-3-2(Conv2D)	(8, 8, 128)	147,584
batchnormalisation-11	(8, 8, 128)	512

Table 1. *Cont.*

Layer (Type)	Output Shape	#Param
upsample-2(UpSampling2D)	(16, 16, 128)	0
dropout-5(Dropout)	(16, 16, 128)	0
upconv-2(Conv2D)	(16, 16, 64)	73,792
batchnormalisation-12	(16, 16, 64)	256
upconv-2-2(Conv2D)	(16, 16, 64)	36,928
batchnormalisation-13	(16, 16, 64)	256
upsample-1(UpSampling2D)	(32, 32, 64)	0
dropout-6(Dropout)	(32, 32, 64)	0
upconv-1(Conv2D)	(32, 32, 32)	18,464
batchnormalisation-14	(32, 32, 32)	128
upconv-1-2(Conv2D)	(32, 32, 32)	9248
batchnormalisation-15	(32, 32, 32)	128
upconv-final(Conv2D)	(32, 32, 3)	867

4. Experimental Results and Discussion

To conduct the experiments, 2642 labelled images captured from 1060 subjects, collected from the Kaggle, YMU, VMU, MIW, MIFS, and FAM datasets, are used to train and evaluate the models. Manual investigation and cleaning are carried out on data. Unclear and low-quality images are removed to reduce their negative impact on supervised models' learning. After reduction, 2479 images are retained, comprising of 1265 image with makeup and 1214 without makeup. To train the unsupervised learning models, 446 unlabelled images are selected from Kaggle dataset [37]. The first step of our system is to detect and locate the face region and crop it. The model trained on the histogram of oriented gradients (HOG) features with support vector machine (SVM) classifier [46] is used to detect the face region. All the detected and cropped faces are then resized into $128 \times 128 \times 3$ and normalised.

For supervised learning, we used pre-trained convolutional neural network VGG16 [42] for feature extraction and fine-tuned it on our datasets. It typically comprises 16 layers including 13 convolution layers of size 3×3 which are followed by five max-pooling layers of size 2×2. After each convolution layer, a rectified linear unit (ReLU) activation function is appended. The probabilities for each category are then generated using softmax layer that follows three dense layers. To apply the fine-tuning using our data, the dense and softmax layers are replaced with new layers. The new layers consist of two fully connected layers with 4096 neurons and a softmax layer. The Stochastic Gradient Descent (SGD) with learning rate of 0.0001 and momentum of 0.9 was used as an optimisation algorithm and binary cross-entropy was used as loss function. Four types of data augmentation were applied during the learning stage to double the size of data artificially. These data augmentation types include horizontal flipping, random zooming, and horizontal and vertical shift. The network was trained for 100 epoch with batch size of 32. The learning curve of VGG16 network during training stage is shown in Figure 6a.

In the second model, the VGG16 trained in the first model is merged with an unsupervised learning scheme via self-learning producing a semi-supervised learning approach. This aims to increase the size of training data by applying self-learning targeting towards producing pseudo-labels. The expansion of training data would help to improve the robustness of the classifier. With samples from the unlabelled dataset that have high confidence when categorised with the VGG16 CNN model, the self-learning approach [23] is exploited to extend the size of the initial training data. Iteratively adding pseudo-labels (obtained

from the self-learning model) to the training dataset is achieved by updating the classifier for each iteration and labelling the rest unlabelled instances. Thus, the appending of pseudo-labels to the training data keeps on until the convergence. The unlabelled examples that remain unclassified at the end of the training stage are discarded due to their unreliability. Finally, the testing images are used to validate the resulted classifier. The same number of epochs, optimiser and learning rate used during VGG16 learning are used here. The learning curve of semi-supervised during the training stage is shown in Figure 6b.

In the third model, unsupervised training using unlabelled data is applied by the convolutional auto-encoder (CAE) model before training the classifier on labelled data, producing a semi-supervised learning scheme. Basically, the auto-encoder tries to reconstruct the original input image using a back-propagation algorithm targeting to reduce the reconstruction error. The goal of using CAE is to find the representative features from the unlabelled data. After training the CAE as an unsupervised method, we remove the decoder components and use the encoder part of CAE for initialising a supervised CNN. On top of these layers, we add two fully connected layers of size 512 and 256, followed by a classification layer (softmax output unit of size two since there are two classes; makeup and without makeup). The weights of the encoder part trained on unlabelled data are used for initialising the CNN training, whereas the weights of fully connected layers are trained from scratch using labelled data. It has been shown that the pre-training of the network allows for higher generalisation performance than when starting from a random weight initialisation. The auto-encoder was trained for 50 epoch with a batch size of 32. Adam optimiser [47] with binary cross-entropy was used for network optimisation with a learning rate of 0.00015. Figure 6c states the learning curve in term of loss for CAE training. To produce the makeup classifier, the encoder part with connected layers are trained on labelled data using an SGD optimiser with a learning rate of 0.001 and momentum of 0.9 for 100 epoch. Figure 6d,e show the system performance during the training in terms of error loss and accuracy, respectively.

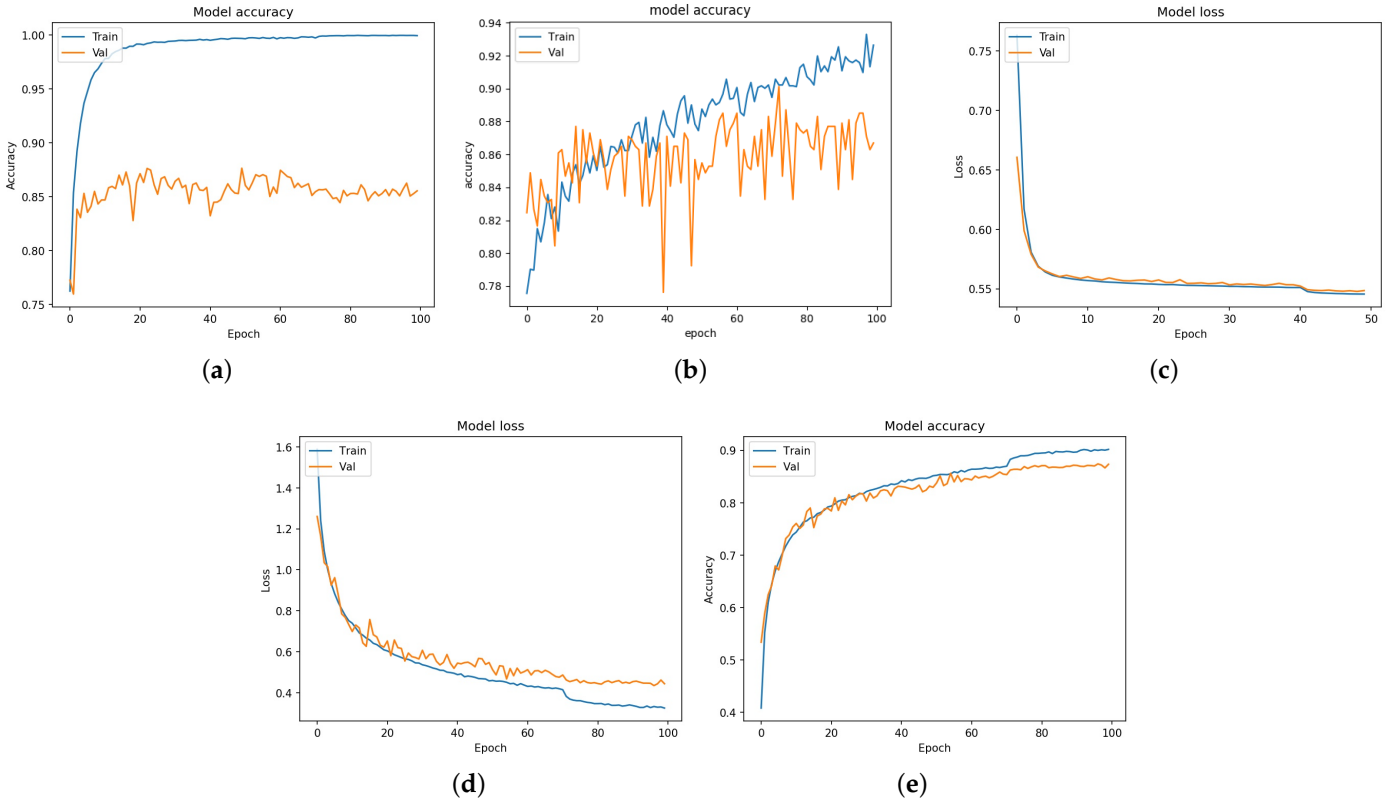
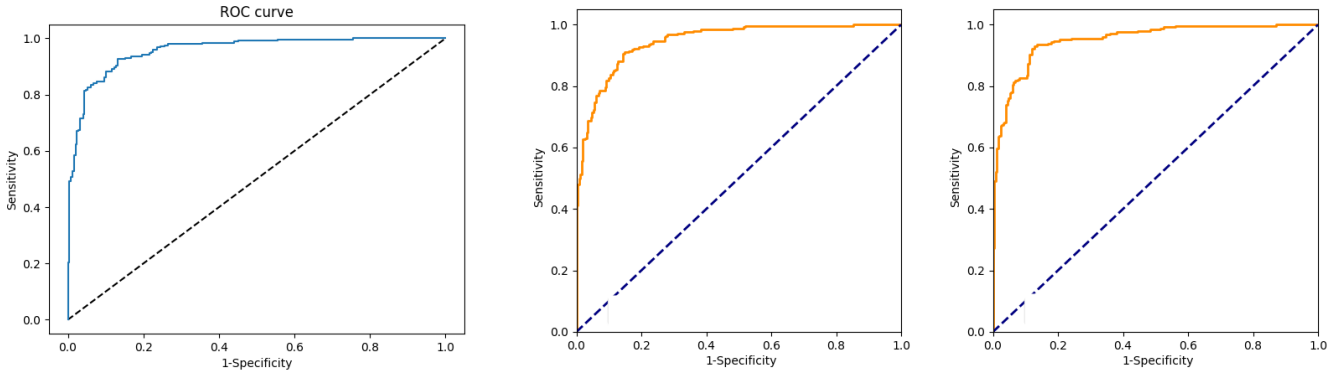


Figure 6. Learning curves of the implemented models (loss and accuracy curves). (**a**) VGG16 (accuracy); (**b**) CNN with Self-learning (accuracy); (**c**) Convolutional auto-encoder (loss); (**d**) Autoencoder-Classifier (loss); (**e**) Autoencoder-Classifier (accuracy).

The proposed makeup detection algorithm's performance is trained and tested using a five-fold cross-validation scheme. In terms of subjects, there is no overlap between the training and testing sets. The five datasets are separated in an equal percentage through the five folds. Four folds are used for learning the makeup detection model, and the remaining fold is considered for model testing. This process is repeated five times to compute the average over the folds in the five-fold cross-validation scheme. The performance of the three models are evaluated using two metrics; accuracy and Area Under Curve of Receiver Operating Characteristic AUROC as reported in Table 2 and displayed in Figure 7.

Table 2. Performance of makeup detection in three models.

Method	Accuracy	AUROC
VGG16 CNN	86.69%	92.30%
CNN with Self-Learning	87.40%	94.69%
Autoencoder-Classifier	88.33%	95.15%



(**a**) (**b**) (**c**)

Figure 7. AUROC curves of the developed models. (**a**) VGG16; (**b**) CNN with Self-learning; (**c**) Autoencoder-Classifier.

The Receiver Operating Characteristic curve, or the ROC curve, is a valuable tool when predicting the probability of a binary classifier. It is a plot of the False Positive Rate FPR (1-specificity) versus the True Positive Rate TPR (sensitivity) for several different candidate threshold values between 0.0 and 1.0. With a ROC curve, the classifier's performance attempts to find a good model that optimises the trade-off between the FPR and TPR. What counts here is how much area is under the curve (AUROC). The ideal curve is when the classifier can distinguish between negative and positive results with 100% area under the curve, which is hardly challenging. An excellent model has AUCROC near the 100%, which means it has a good separability measure. A poor model has an AUCROC near 0%, which means it has the worst separability measure. The shape of the curve contains a lot of information, including the predicted FPR and the TPR. Smaller values on the x-axis of the plot indicate lower false positives and higher true negatives. Likewise, larger values on the y-axis of the plot indicate higher true positives and lower false negatives. The models understudy attained AUROC of 92.30%, 94.69%, and 95.15% in VGG16CNN, CNN with Self-Learning, and Autoencoder-Classifier, respectively. The obtained results reveal high detection outcomes in all classification models represented by curves that bow up to the top left of the plot. This also concludes that the classifier can detect more numbers of true positives and true negatives than false negatives and false positives.

We conducted a thorough investigation using five publicly available labelled datasets. This helps us compare our methods to earlier work by evaluating image datasets collected from various sources, demonstrating the reliability of our methods. When it comes to comparing our suggested methods to other approaches in the literature, we carry out a

direct comparison with work presented in [31] who tested their system on the same datasets we used in our experiments and used a similar data setup. Their method achieved an accuracy of 79% using a deep supervised learning approach which is apparently lower than our obtained performance. Wang et al. [5] have evaluated their methods on YMU, MIW, and VMU, achieving an accuracy of 91.59%, 91.41%, and 93.75%, respectively. Moreover, Chen et al. [25] reported accuracy of 89.94% on the YMU dataset. However, both research works used features extracted manually, which might be applicable to the dataset they have used but not to other datasets. Furthermore, Wang et al. [5] pre-processed and aligned all images in the different datasets using facial landmarks. This makes the performance of their method is highly affected and biased by the alignment process. Whereas our method requires no pre-processing step except for image normalisation. We also do not know which subjects were picked up and allocated for training and testing their models, making the comparison not exact. The authors in [24,26,29,32,39] are reported face verification and vulnerability assessment performance but not facial makeup detection performance. In our work, we did not conduct face recognition experiments as our methodology's objective is only to detect the presence of the facial makeup. Moreover, our method outperformed the makeup detection method presented in [4] that achieved detection accuracy of 55.5%, 52%, 72.5%, and 52.5% using colour, smoothness, texture, and highlight features, respectively.

5. Conclusions

In this work, we conducted an investigation relating to makeup detection using multiple learning schemes. Three learning mechanisms, including supervised, unsupervised, and semi-supervised with transfer learning approaches, have been exploited to detect the presence of makeup in the facial images as a binary classification task. The obtained experimental results demonstrated that our methods efficiently identify the makeup and generalise well on unseen data compared to the existing makeup detection methods. It also reveals that unlabelled data can yield changes in learning accuracy when used in combination with a limited amount of labelled data. Thus, semi-supervised learning can be of great practical benefit in such cases, which improves the robustness of the developed models. Considering these findings, in the future, we propose to extend our makeup detection system by developing a makeup removal algorithm that depends on the makeup detection performance. The makeup removal system could recover the bare face, which helps improving facial recognition frameworks and other facial-related systems. Moreover, other pre-trained CNN models, including EfficientNet, InceptionV3, DenseNet, etc., could be further investigated, and their performance could be studied and benchmarked comparing to VGG16 in future work.

Author Contributions: T.A. conceived and designed the experiments; T.A. and B.A.-B. performed the experiments; T.A., B.A.-B. and W.A.-N. analysed the data and wrote the paper; W.A.-N. supervised and ran the project. All authors have read and agreed to the published version of the manuscript.

Funding: Theiab Alzahrani was funded by the Kingdom of Saudi Arabia government.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analysed in this study. The reference of each dataset has been cited in the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DL	Deep Learning
CNN	Convolutional Neural Networks
YMU	YouTube Makeup
VMU	Virtual Makeup
MIW	Makeup in the Wild
MIFS	Makeup Induced Face Spoofing
FAM	FAce Makeup
CAE	Convolutioanl Auto-Encoder

References

1. Guéguen, N. The effects of women's cosmetics on men's courtship behaviour. *N. Am. J. Psychol.* **2008**, *10*, 221–228.
2. Dantcheva, A.; Dugelay, J. Female facial aesthetics based on soft biometrics and photo-quality. In Proceedings of the 2011 IEEE International Conference on Multimedia and Expo (ICME'11), Washington, DC, USA, 11–15 July 2011; Volume 2.
3. Liu, X.; Li, T.; Peng, H.; Chuoying Ouyang, I.; Kim, T.; Wang, R. Understanding beauty via deep facial features. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019; pp. 246–256.
4. Guo, G.; Wen, L.; Yan, S. Face authentication with makeup changes. *IEEE Trans. Circuits Syst. Video Technol.* **2013**, *24*, 814–825.
5. Wang, S.; Fu, Y. Face behind makeup. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; Volume 30.
6. Rathgeb, C.; Dantcheva, A.; Busch, C. Impact and detection of facial beautification in face recognition: An overview. *IEEE Access* **2019**, *7*, 152667–152678. [CrossRef]
7. Rathgeb, C.; Satnoianu, C.I.; Haryanto, N.; Bernardo, K.; Busch, C. Differential detection of facial retouching: A multi-biometric approach. *IEEE Access* **2020**, *8*, 106373–106385. [CrossRef]
8. Sajid, M.; Ali, N.; Dar, S.H.; Iqbal Ratyal, N.; Butt, A.R.; Zafar, B.; Shafique, T.; Baig, M.J.A.; Riaz, I.; Baig, S. Data augmentation-assisted makeup-invariant face recognition. *Math. Probl. Eng.* **2018**, *2018*, 2850632. [CrossRef]
9. Dong, X.; Yan, Y.; Ouyang, W.; Yang, Y. Style aggregated network for facial landmark detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 379–388.
10. Li, Y.; Song, L.; Wu, X.; He, R.; Tan, T. Anti-makeup: Learning a bi-level adversarial network for makeup-invariant face verification. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; Volume 32.
11. Li, Y.; Song, L.; Wu, X.; He, R.; Tan, T. Learning a bi-level adversarial network with global and local perception for makeup-invariant face verification. *Pattern Recognit.* **2019**, *90*, 99–108. [CrossRef]
12. Shen, X. Facebook and Huawei Are the Latest Companies Trying to Fool Facial Recognition. 2019. Available online: https://www.scmp.com/abacus/tech/article/3035451/facebook-and-huawei-are-latest-companies-trying-fool-facial-recognition (accessed on 27 September 2021).
13. Valenti, L. Can Makeup Be an Anti-Surveillance Tool? 2020. Available online: https://www.vogue.com/article/anti-surveillance-makeup-cv-dazzle-protest (accessed on 27 September 2021).
14. Valenti, L. Yes, There Is a Way to Outsmart Facial Recognition Technology—And It Comes Down to Your Makeup. 2018. Available online: https://www.vogue.com/article/computer-vision-dazzle-anti-surveillance-facial-recognition-technology-moma-ps1 (accessed on 27 September 2021).
15. Gafni, O.; Wolf, L.; Taigman, Y. Live face de-identification in video. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 9378–9387.
16. Alzahrani, T.; Al-Nuaimy, W. Face segmentation based object localisation with deep learning from unconstrained images. In Proceedings of the 10th International Conference on Pattern Recognition Systems (ICPRS-2019), Tours, France, 8–10 July 2019.
17. Alzahrani, T.; Al-Nuaimy, W.; Al-Bander, B. Hybrid feature learning and engineering based approach for face shape classification. In Proceedings of the 2019 International Conference on Intelligent Systems and Advanced Computing Sciences (ISACS), Taza, Morocco, 26–27 December 2019; pp. 1–4.
18. Alzahrani, T.; Al-Nuaimy, W.; Al-Bander, B. Integrated Multi-Model Face Shape and Eye Attributes Identification for Hair Style and Eyelashes Recommendation. *Computation* **2021**, *9*, 54. [CrossRef]
19. Vincent, P.; Larochelle, H.; Bengio, Y.; Manzagol, P.A. Extracting and composing robust features with denoising autoencoders. In Proceedings of the 25th International Conference on Machine Learning, New York, NY, USA, 5–9 July 2008; pp. 1096–1103.
20. Masci, J.; Meier, U.; Cireşan, D.; Schmidhuber, J. Stacked convolutional auto-encoders for hierarchical feature extraction. In Proceedings of the International Conference on Artificial Neural Networks, Espoo, Finland, 14–17 June 2011; pp. 52–59.
21. Kingma, D.P.; Mohamed, S.; Jimenez Rezende, D.; Welling, M. Semi-supervised learning with deep generative models. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 3581–3589.
22. Odena, A. Semi-supervised learning with generative adversarial networks. *arXiv* **2016**, arXiv:1606.01583.

23. Lee, D.H. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In Proceedings of the Workshop on Challenges in Representation Learning, ICML, Daegu, Korea, 3–7 November 2013; Volume 3.
24. Dantcheva, A.; Chen, C.; Ross, A. Can facial cosmetics affect the matching accuracy of face recognition systems? In Proceedings of the 2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS), Arlington, VA, USA, 23–27 September 2012; pp. 391–398.
25. Chen, C.; Dantcheva, A.; Ross, A. Automatic facial makeup detection with application in face recognition. In Proceedings of the 2013 International Conference on Biometrics (ICB), Madrid, Spain, 4–7 June 2013; pp. 1–8.
26. Hu, J.; Ge, Y.; Lu, J.; Feng, X. Makeup-robust face verification. In Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, Canada, 26–31 May 2013; pp. 2342–2346.
27. Dantcheva, A.; Dugelay, J.L. Assessment of female facial beauty based on anthropometric, non-permanent and acquisition characteristics. *Multimed. Tools Appl.* **2015**, *74*, 11331–11355. [CrossRef]
28. Bharati, A.; Singh, R.; Vatsa, M.; Bowyer, K.W. Detecting facial retouching using supervised deep learning. *IEEE Trans. Inf. Forensics Secur.* **2016**, *11*, 1903–1913. [CrossRef]
29. Kamil, I.A.; Are, A.S. Makeup-invariant face identification and verification using fisher linear discriminant analysis-based Gabor filter bank and histogram of oriented gradients. *Int. J. Signal Imaging Syst. Eng.* **2017**, *10*, 257–270. [CrossRef]
30. Fu, Y.; Shuyang, W. System for Beauty, Cosmetic, and Fashion Analysis. U.S. Patent 10,339,685, 2 July 2019. Available online: https://patents.google.com/patent/US20170076474A1/en (accessed on 13 October 2021).
31. Li, M.; Li, Y.; He, Y. *Makeup Removal System with Deep Learning*; Technical Report, CS230 Deep Learning; Stanford University: Stanford, CA, USA, 2018.
32. Rathgeb, C.; Drozdowski, P.; Fischer, D.; Busch, C. Vulnerability assessment and detection of makeup presentation attacks. In Proceedings of the 2020 8th International Workshop on Biometrics and Forensics (IWBF), Porto, Portugal, 29–30 April 2020; pp. 1–6.
33. Rathgeb, C.; Botaljov, A.; Stockhardt, F.; Isadskiy, S.; Debiasi, L.; Uhl, A.; Busch, C. PRNU-based detection of facial retouching. *IET Biom.* **2020**, *9*, 154–164. [CrossRef]
34. Rathgeb, C.; Drozdowski, P.; Busch, C. Makeup Presentation Attacks: Review and Detection Performance Benchmark. *IEEE Access* **2020**, *8*, 224958–224973. [CrossRef]
35. Fu, Y.; Guo, G.; Huang, T.S. Age synthesis and estimation via faces: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1955–1976. [PubMed]
36. Kotwal, K.; Mostaani, Z.; Marcel, S. Detection of age-induced makeup attacks on face recognition systems using multi-layer deep features. *IEEE Trans. Biom. Behav. Identity Sci.* **2019**, *2*, 15–25. [CrossRef]
37. Kaggle. Makeup or No Makeup. 2018. Available online: https://www.kaggle.com/petersunga/make-up-vs-no-make-up (accessed on 1 November 2020).
38. Chen, C.; Dantcheva, A.; Ross, A. An ensemble of patch-based subspaces for makeup-robust face recognition. *Inf. Fusion* **2016**, *32*, 80–92. [CrossRef]
39. Chen, C.; Dantcheva, A.; Swearingen, T.; Ross, A. Spoofing faces using makeup: An investigative study. In Proceedings of the 2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA), New Delhi, India, 22–24 February 2017; pp. 1–8.
40. NIST. Face Recognition Grand Challenge (FRGC). 2018. Available online: http://www.nist.gov/itl/iad/ig/frgc.cfm (accessed on 1 November 2020).
41. Taaz. TAAZ Makeover Technology. Available online: http://www.taaz.com/ (accessed on 1 November 2020).
42. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
43. Torrey, L.; Shavlik, J. Transfer learning. In *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques*; IGI Global: Pennsylvania, PA, USA, 2010; pp. 242–264.
44. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
45. Bengio, Y.; Lamblin, P.; Popovici, D.; Larochelle, H. Greedy layer-wise training of deep networks. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 3–6 December 2007; pp. 153–160.
46. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the International Conference on Computer Vision & Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893.
47. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.