# Improvement of Facial Beauty Prediction Using Artificial Human Faces Generated by Generative Adversarial Network

Donatas Laurinavičius[1] · Rytis Maskeliūnas[1] · Robertas Damaševičius[2]

## Abstract

Human beauty evaluation is a particularly difficult task. This task can be solved using deep learning methods. We propose a new method for determining the attractiveness of a face by using the generation of synthetic data. Our approach uses the generative adversarial network (GAN) to generate an artificial face and then predict the facial beauty of the generated face to improve facial beauty predictions. A study of images with different brightness and contrast showed that the methods using the convolutional neural network (CNN) model have fewer errors than compared to the multilayer perceptron (MLP) model that uses the method. The MLP model only responds to geometric facial proportions, whereas the CNN model additionally responds to changes in face color. Using the synthetic face instead of the real face improves the determination of accuracy of the facial attractiveness. The ability to appreciate facial beauty also opens the way for facial beauty modifications in a latent space. Further research could improve facial normalization in the latent space to improve the accuracy of facial beauty determination.

**Keywords** Facial beauty assessment · Human face generation · Generative adversarial networks

## Introduction

Beauty is in the eye of the beholder. Because of humans' unusually well-developed ability to interpret, identify, and extract information from other people's features, the human face has piqued the interest of psychologists and other scientists in recent years. Our publications and television screens are not simply loaded with faces; they are filled with lovely faces, and both men and women are anxious about a possible partner's appearance. Humans value their physical appearance, and some characteristics appear to be desirable across people and cultures [1, 2]. Cunningham et al.'s multidimensional fitness model of physical beauty proposes that perception of high physical attractiveness incorporates a number of desirable traits and personal attributes [3]. Such characteristics might be assessed using a biologically inspired face analysis system. To assess a face without undertaking a blind spatial search, terms such as saliency or gist may be employed. The

saliency model is made up of highly concurrent low-level calculations in domains including intensity, direction, and hue. It is used as a starting point to draw attention to a group of prominent spots in a picture. When used with the saliency model, the gist model may offer predicted holistic image attributes [4]. Empirical evidence suggests that there is an optimal arrangement of facial characteristics (ideal ratios) that can improve a person's face's beauty [5]. Computational prediction of face attractiveness has gained significant scientific attention, with several applications in multimedia. Bio-inspired, deep learning–based discriminative representations for face aesthetic prediction can aid in identifying needed spatial regions of interest during human subjects' facial aesthetic evaluations [6], leading towards the motivation of our research, that is the dealing with the main difficulty in such approaches: to extract discriminative and perception-aware elements that can be used to describe facial beauty.

The perception of human beauty is naturally subjective, but when assessing the beauty and attractiveness of other people, it is often accepted if the person being evaluated meets certain beauty standards or, openly, does not meet them. This agreement on who is and is not beautiful is so widespread that competitions to assess human beauty are held [7, 8]. One of the best-known examples where beauty is the key criterion is the Miss Universe competition [9]. The

✉ Robertas Damaševičius
robertas.damasevicius@polsl.pl

1 Faculty of Informatics, Multimedia Engineering Department, Kaunas University of Technology, Kaunas, Lithuania

2 Faculty of Applied Mathematics, Silesian University of Technology, Gliwice, Poland

long history and popularity of such competitions imply the existence of beauty and attractiveness criterion that most people agree on.

Facial beauty analysis [10–12] is used for a variety of purposes, including face enhancement programs (MeiTu, FaceTune) and plastic surgery. Some diseases, such as facial palsy, can be diagnosed by analyzing facial characteristics such as symmetry [13]. To perform this analysis, methods for assessing the beauty and attractiveness of the human face have already been developed, including facial proportions [14], the golden ratio [15], ideal dimensions [16], and geometric features [17]. Although facial beauty prediction (FBP) has achieved high accuracy in photos taken in a controlled environment, it remains a challenging problem in real-world face photographs [18]. Assessing human beauty is a particularly difficult task, as it is affected by numerous variables such as photo resolution, human face angle, and lighting [19]. Artificial intelligence methods are frequently used to solve such problems [20–22].

Human face images are now commonly analyzed using machine learning and computer vision techniques. The human facial image conveys information such as age, gender, identity, emotion, race, and attractiveness to both humans and computers [23]. Several studies on facial attractiveness have been conducted. The perception of facial attractiveness is highly subjective and can be influenced by sociological or cultural factors as well as personal desires. Iyer et al. [24], for example, extracted facial landmarks to create facial ratios based on Golden and Symmetry Ratios. In the Hue, Saturation, and Value (HSV) space, texture, shape, and color features are retrieved as Gray Level Covariance Matrix (GLCM), Hu's Moments, and Color Histograms, respectively. Another ablation trial is performed to determine which feature, when combined with facial landmarks, works best. In experiments, combining key facial traits with facial landmarks increased the facial beauty prediction score. Some facial features are objectively more appealing than others [25]. Objects that have a ratio are considered harmonious and beautiful [26]. Face studies show that symmetrical faces are more appealing [27]. In the case of identical twins, one study found that a twin with more symmetric proportions was considered more attractive [28]. The uniformity of people's faces is defined by their mass as well as their similarity to other people. Different people's faces have distinct facial features that set them apart from the majority of the population. Some research suggests that identical faces are more attractive [29]. People with identical faces are more likely to be symmetrical, and symmetrical faces, as previously observed, are considered more attractive. Baby-like characteristics are linked to sympathy and people's proclivity to patronize protection. A large, rounded forehead, low position of the eyes and mouth, large, round eyes, and a low chin distinguish baby features. The study of youthful faces also shows that face attractiveness is positively related to its youthfulness [29].

According to research [30], there is a link between human facial health and facial attractiveness. One of the indicators of good human health is healthy facial skin. Furthermore, studies show that people are more likely to associate skin redness with being in good health. Human attractiveness is directly related to human wellness [27]. According to these studies, certain facial proportions are objectively more appealing to the majority of people [19]. A popular method for determining the accuracy of beauty determination, in which the Pearson correlation coefficient [31] is computed. It allows evaluating how strongly the method of determining beauty correlates with the human beauty determined in the opinion of a real expert. Ideally, the method of determining beauty should have a Pearson correlation of 1.

Another way to determine the beauty and attractiveness of the face is considered more modern and is becoming more and more popular, based on deep learning [32–35]. Convolutional neural networks (CNN) and other deep learning models can be used to automatically recognize facial features that determine the beauty and attractiveness of the face. This method is used by researchers to assess the beauty and attractiveness of the human face. For example, ResNet50, one of the more advanced architectures in convoluted networks, was used in the study [34]. The researchers claim to have achieved Pearson's correlation coefficient of 0.87 after training the network with their dataset.

Vahdati et al. [35] employ a multi-task learning strategy to identify the best shared features for three related tasks (i.e., facial beauty assessment, gender recognition, and ethnicity identification). To improve attractiveness calculation accuracy, specific parts of face images (e.g., the left eye, nose, and mouth) as well as the entire face are fed into multi-stream CNNs. Each two-stream network accepts both a portion of the face and the entire face as input. Beauty3DFaceNet, the first deep learning network for evaluating attractiveness in 3D faces, is proposed by Xiao et al. [36]. It combines facial geometry, texture, and history to produce a more realistic 3D facial attractiveness score, similar to that of human raters. They also provide 3DFacePointNet++, a novel network based on facial landmark priors that improves the Beauty3DFaceNet's performance by simulating human eye perceptual sensitivity.

Lin et al. [21] define facial beauty prediction as a special regression problem driven by ranking data. We present R3CNN, a general CNN architecture that incorporates the relative ranking of faces in terms of aesthetics, to improve the performance of Facial beauty prediction. Bougourzi et al. [37] propose a two-branch architecture (REX-INCEP) based on merging the architecture of two already trained networks to deal with the difficult high-level features associated with the facial beauty prediction problem. They present an ensemble regression method based on CNNs and employ

both networks in this ensemble (REX-INCEP). Recently, generative adversarial networks (GANs) [38], a type of deep generative neural architecture, have enabled unprecedented realism in the generation of synthetic human faces [39], landscapes and buildings [40], and medical images [41]. This paper aims to improve human beauty and attractiveness by using GANs to predict human beauty and attractiveness.

This paper's contribution is an innovative GAN-based methodology for predicting human beauty and attractiveness. Our bio-inspired approach allows using a generated face instead of the real face to enhance the accuracy of determining facial attractiveness. The remainder of the paper is structured as follows: the "Method" section focuses on the technique and algorithms developed, while the "Experiments and Results" section discusses experimental assessment and findings. The

"Discussion and Conclusions" section concludes the article and discusses our future research plans.

## Method

### Outline of the Methodology

Figure 1 depicts an outline of the methodology. The suggested approach has the following stages described further in the article: (1) detection of faces in photos; (2) generation of an artificial face for each detected face; (3) extraction of facial characteristics from a generated face; (4) evaluation of each face using a Multilayer Perceptron (MLP) model
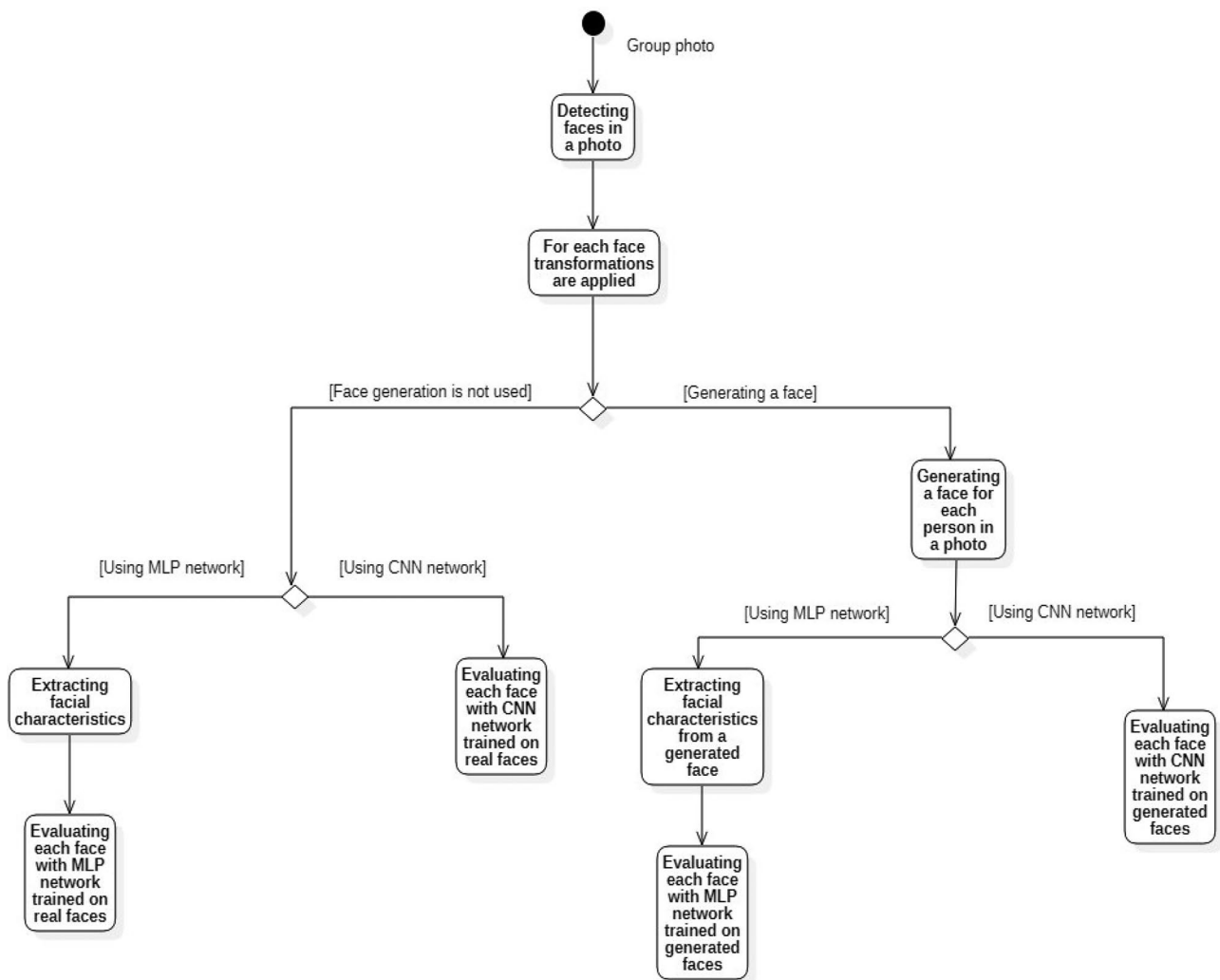


**Fig. 1** Outline of the methodology

trained on generated faces; (5) evaluation of each face using a CNN model trained on generated faces. The findings are then compared to the results of face evaluation using MLP and CNN but without the use of artificial face creation.

## Face Detection in Group Photos

Recognizing people's faces in group photographs was the initial step towards analyzing them. A histogram of directed gradients was used to detect facial characteristics (HOG) and the support vector machine (SVM) was then trained for face identification using features found in 3000 photos from the Labeled Faces in the Wild dataset [42].

## Extracting Facial Characteristics

The identified face then underwent required transformations in order to prepare for attractiveness analysis. These changes allowed considerable increase in the accuracy of assessing face beauty. The 81 face points were retrieved using the "dlib" package and a custom-trained model [43]. Once the mesh vector was in place, the face underneath it was rotated, cut off, and turned into a $1024 \times 1024$ pixel picture.

Before extracting facial characteristics, facial transformations were performed, and then the necessary facial points were extracted for each transformed face, or if the points needed to determine the attractiveness of the face using the generated facial copy are extracted for the generated face, as summarized in Table 1, which employs the face characteristic measurements shown in Fig. 2.
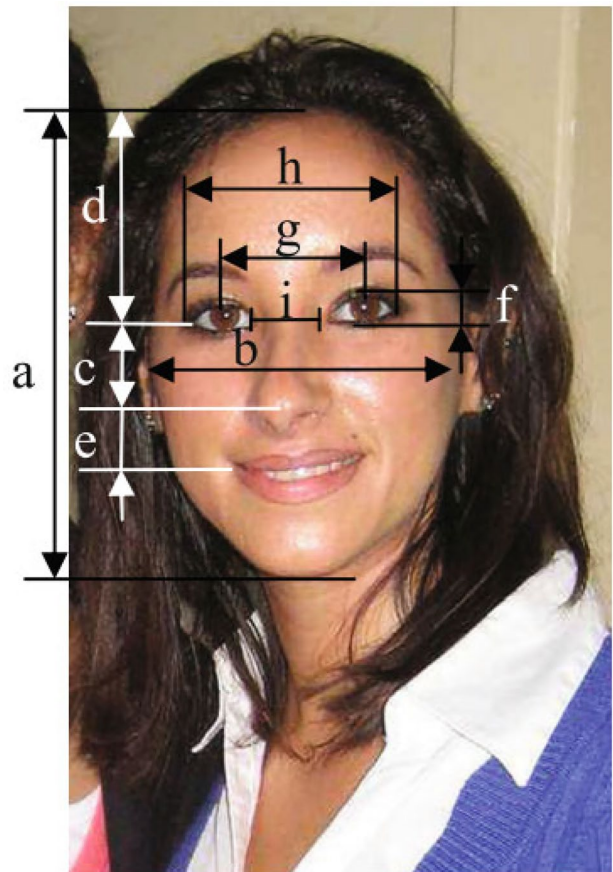


**Fig. 2** Measurements of facial characteristics: a – face length, b – face width, c – distance from eye to nose, d– distance from the top to the eye, e – distance from nose to lips, f – eye height, g – distance between eyes, h – distance between outer corners of eyes, i – distance between inner corners of eyes

**Table 1** Facial characteristics used to assess facial attractiveness

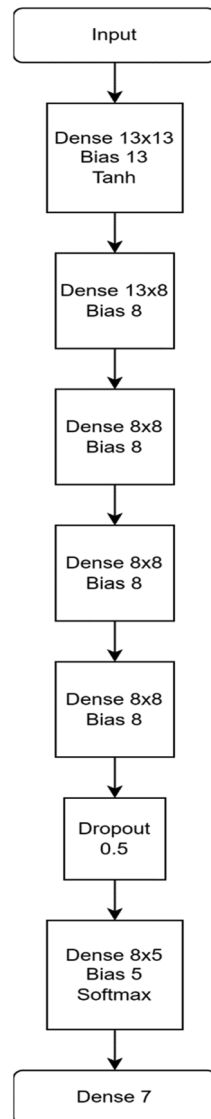| No. | Characteristic | Calculation |
|---|---|---|
| 1 | Eye height/face length | f/a |
| 2. | Face width/face length | b/a |
| 3. | Top to nose/face length | (d + c) / a |
| 4. | Top to mouth/face length | (d + c + e) / a |
| 5. | Top to eye/face length | d/a |
| 6. | Eye width/eye gap | (h – i) / 2i |
| 7. | Nose to chin/eye to nose | (a – d – c) / c |
| 8. | From the top to the eye/from the eye to the nose | d / c |
| 9. | Gap between outer corners of the eyes/from the top to the eye | h / d |
| 10. | From eye to nose/nose to mouth | c / e |
| 11. | From the top to the eye/from the eye to the lip | d / (c + e) |
| 12. | Eye to nose/facial width | c/b |
| 13. | Nose to chin/from lip to chin | (a-d-c) / (a-d-c-e) |

## Multilayer Perceptron Model

A MLP model was used to determine the attractiveness of the face using facial characteristics. Network input had 13 facial characteristics, while the network output indicated a probability for each possible estimate (Fig. 3). Each reflected a possible assessment of attractiveness.

A modified ResNet50 model, customized to compute the likelihood of each estimate, was used to measure face attractiveness using convoluted networks, similarly to the MLP model above. Instead of face attributes, the photo itself was fed as input. The model was trained to detect the facial traits that define the beauty of the face and to score the attractiveness of the photo based on these aspects. Figure 4 depicts the architecture of the CNN model.

GAN (generative adversarial network) model consists of two parts: a generator(s). *generator*) and discriminator and

**Fig. 3** MLP network architecture CNN model

```
┌─────────────────┐
│      Input      │
└─────────────────┘
         ↓
┌─────────────────┐
│  Dense 13x13    │
│    Bias 13      │
│      Tanh       │
└─────────────────┘
         ↓
┌─────────────────┐
│   Dense 13x8    │
│     Bias 8      │
└─────────────────┘
         ↓
┌─────────────────┐
│   Dense 8x8     │
│     Bias 8      │
└─────────────────┘
         ↓
┌─────────────────┐
│   Dense 8x8     │
│     Bias 8      │
└─────────────────┘
         ↓
┌─────────────────┐
│   Dense 8x8     │
│     Bias 8      │
└─────────────────┘
         ↓
┌─────────────────┐
│    Dropout      │
│      0.5        │
└─────────────────┘
         ↓
┌─────────────────┐
│   Dense 8x5     │
│     Bias 5      │
│    Softmax      │
└─────────────────┘
         ↓
┌─────────────────┐
│    Dense 7      │
└─────────────────┘
```

discriminator. The operation of the generator and the discriminator can also be expressed in the formula (see Eq. (1)) which is the price (*cost*) function $V(G, D)$:

$$\max_G \min_D V(D, G) = E_{x \sim P(x)}[\log(D(x))] + E_{z \sim P(z)}[\log(1 - D(G(z)))] \quad (1)$$

Here, $G$ is the generative model (generator), $D$ is the discriminator, $z$ is noise, $P$ is distribution function, and $x$ is input.

Face generation uses the StyleGAN2-Ada [44] network. To generate a copy of the real face using the GAN network, our approach moves a real face from the photo to *latent* space.

Face generation in *latent* space is shown in Fig. 5. The encoder was used to reduce the amount of image data. The decoder was used to recover the encoded image. The decoder extracts the image from the compressed secret space from the encoder and returns it to its original state before encoding.

We have used a VGG16 model to extract and store face characteristics in a latent space. The generator then returns the encoded face in latent space to its original form. During the facial design process, a vector in a latent space is sought that, if decoded, would restore the papered face. The scenario below depicts the design of a face in a latent space.

Face generation in *latent* space is performed as follows:

1 The generator generates a face for the latent space for the vector
2 The face generated by the generator is placed in a *latent* space using the VGG16 network
3 The real face is placed in a *latent* space using the VGG16 network
4 Calculates the error between the generated face and the face in the secret *latent*) in space
5 Optimizes to reduce error

The procedure is summarized as an algorithm in Fig. 6.

## Experiments and Results

This section describes several experiments that were carried out in order to increase the accuracy of face beauty rating under diverse settings. Faces were examined under non-standard settings, such as poor resolution, in each experiment. The goal was to reduce mean absolute error (MAE). Also, beauty estimate (Mean Opinion Scale (MOS)) is presented for a better understanding of how certain cases lead to the determination of beauty rating [45].

### Training and Evaluation of the MLP Network

We have generated a duplicate of each of the faces in SCUT-FBP5500 dataset using the Style GAN2-ADA network, before training the network. Characteristics for each face were produced after recognizing people's faces in a group shot and completing the appropriate changes, and these characteristics are transferred to the MLP network. The network outputted the probability of each estimate, which was used to calculate the MOS (actual attractiveness rating). Figure 7 depicts network training using stochastic gradient descent. Following the generation of face clones, training was carried out utilizing generated images and attractiveness ratings for original faces. After training, the network achieved a Pearson correlation of 0.741. The training outcomes are shown in Table 2. The following parameters are used to apply the stochastic gradient landing: learning frequency - 0.001, descent - $1 * 10^{-6}$, moment - 0.9. The training was carried out for 100 epochs.

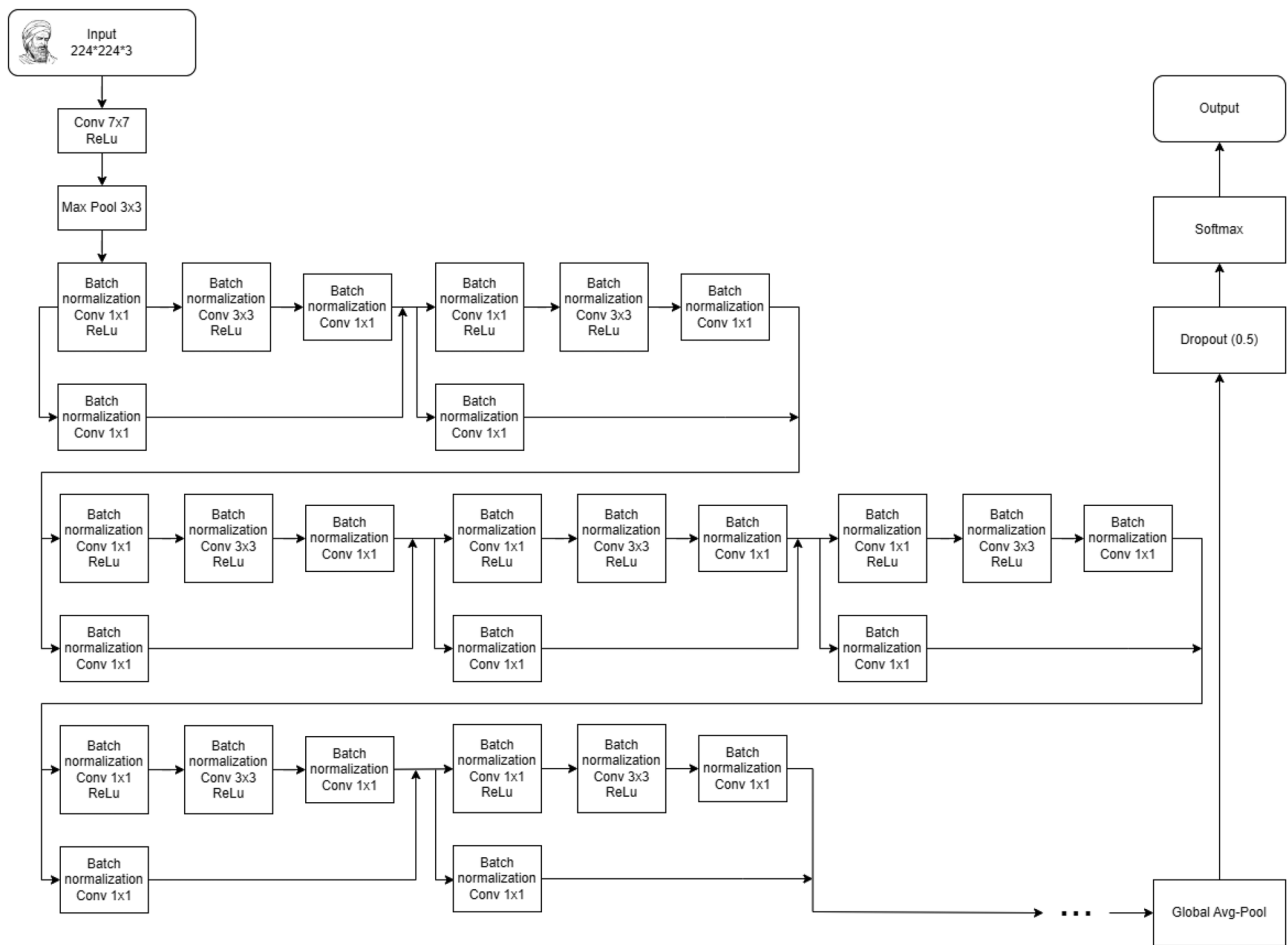Figure 8 depicts the ROC curves of the MLP model with 250 faces taken from the SCUT-FBP5500 dataset for testing with a

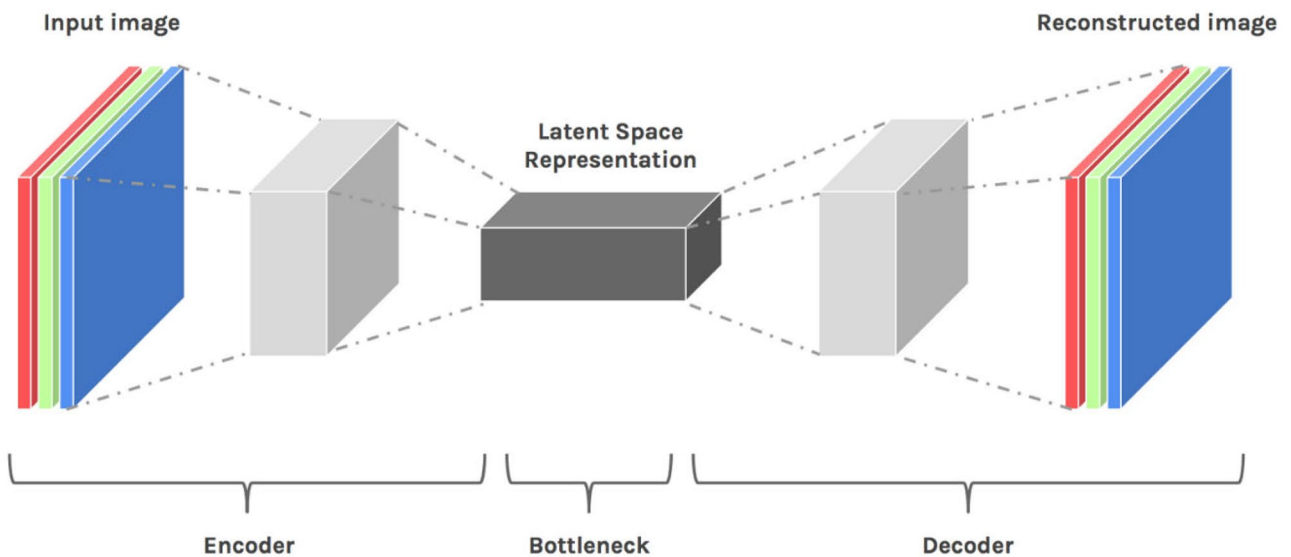**Fig. 4** CNN model architecture face generation using GAN model



**Fig. 5** Encoding in a GAN image puts it in a latent space and restores the image from latent space

**Fig. 6** Algorithm of synthetic face generation

```
Algorithm 1: Face generation in latent space
Data: Input vector X, generator G, network N, real face image F_R
Result: Synthetic face image F_S
while Error > eps do
    /* Repeat while Error is larger than threshold        */
    F_S = G-> generate_face(X) /* Generate synthetic face   */
    L_S = N(F_S) /* Embedding of synthetic face in latent space  */
    L_R = N(F_R) /* Embedding of real face in latent space       */
    Error = diff(L_S, L_R) /* Difference between synthetic and real face */
end
```

**Fig. 7** MLP network training using generated faces
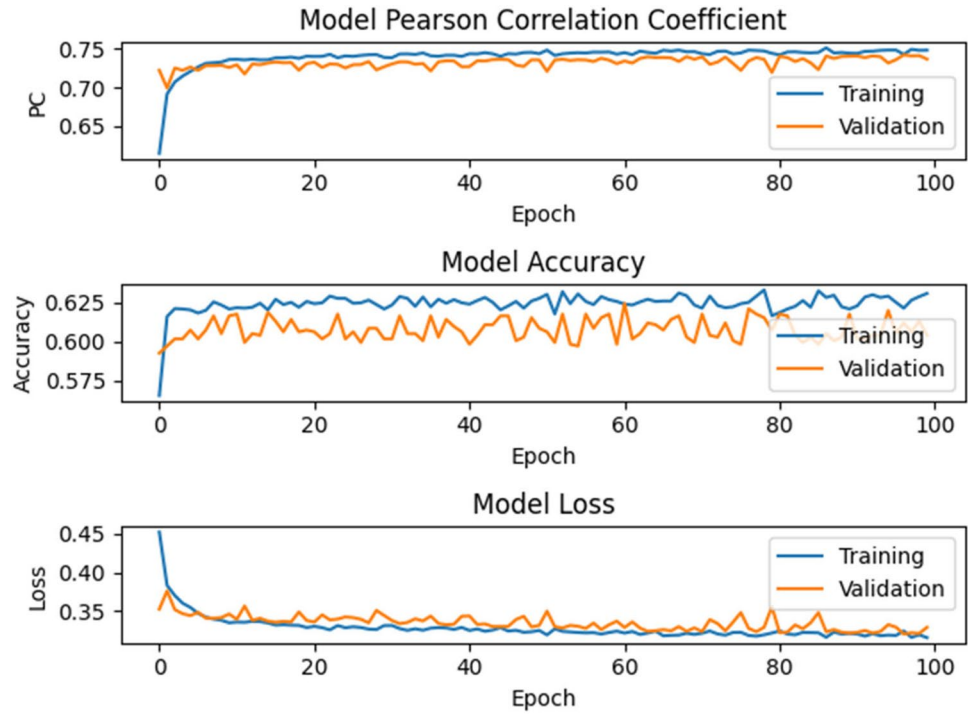


goal to compare the accuracy of beauty evaluation with a wide variation of facial attractiveness. For each facial beauty assessment, 50 faces were chosen from the data set with attractiveness ratings of 1 to 5. Figure 9 indicates that the inaccuracy in the facial attractiveness assessment is similarly wrong with all

conceivable estimations of facial attractiveness, since the curves are next to each other. The MLP model enhances the area under the average ROC curve from 0.76 to 0.78 by using face creation. The larger area under the ROC curve indicates that the attractiveness of the face is determined with greater precision.

**Table 2** MLP network training results using generated faces

| Faces | Number of images for learning | Number of images for testing | Loss | Accuracy | Pearson's correlation coefficient | Mae | RMSE |
|---|---|---|---|---|---|---|---|
| Asian women | 1602 | 398 | 0.323 | 0.565 | 0.724 | 0.118 | 0.159 |
| Asian men | 1601 | 399 | 0.297 | 0.589 | 0.759 | 0.107 | 0.151 |
| Caucasian women | 601 | 149 | 0.396 | 0.570 | 0.677 | 0.123 | 0.167 |
| Caucasian men | 601 | 149 | 0.311 | 0.671 | 0.774 | 0.108 | 0.151 |
| **Total** | **4405** | **1095** | **0.322** | **0.589** | **0.741** | **0.114** | **0.156** |

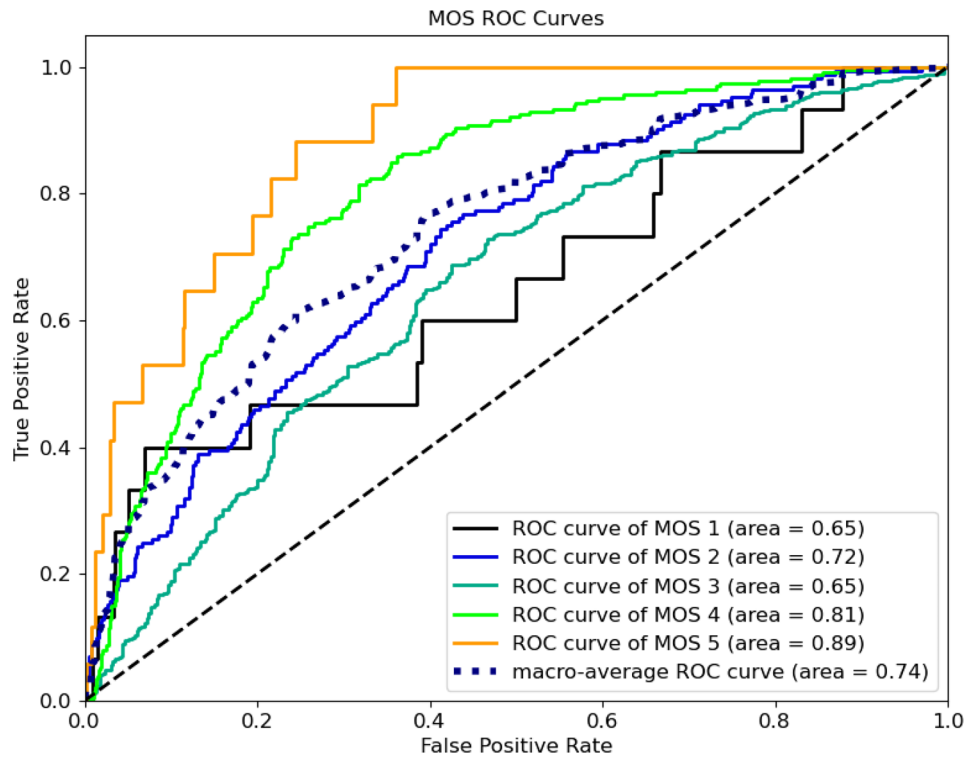**Fig. 8** MLP network ROC curves using generated faces



**Fig. 9** MLP network ROC curves using generated faces with a sample in which the face attractiveness estimate (MOS) is evenly distributed
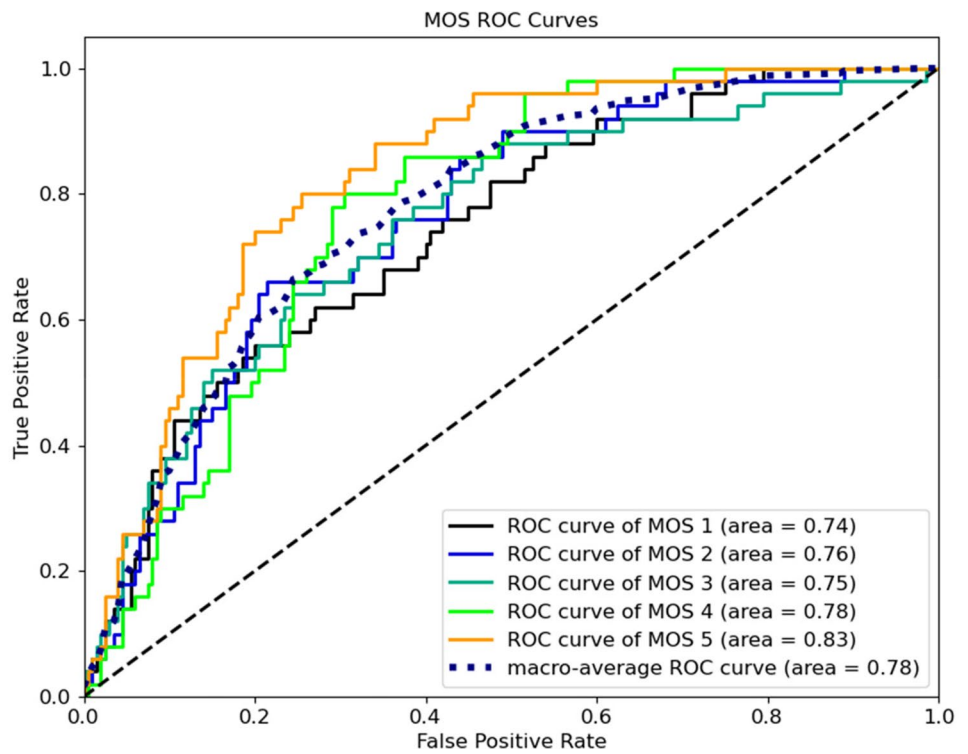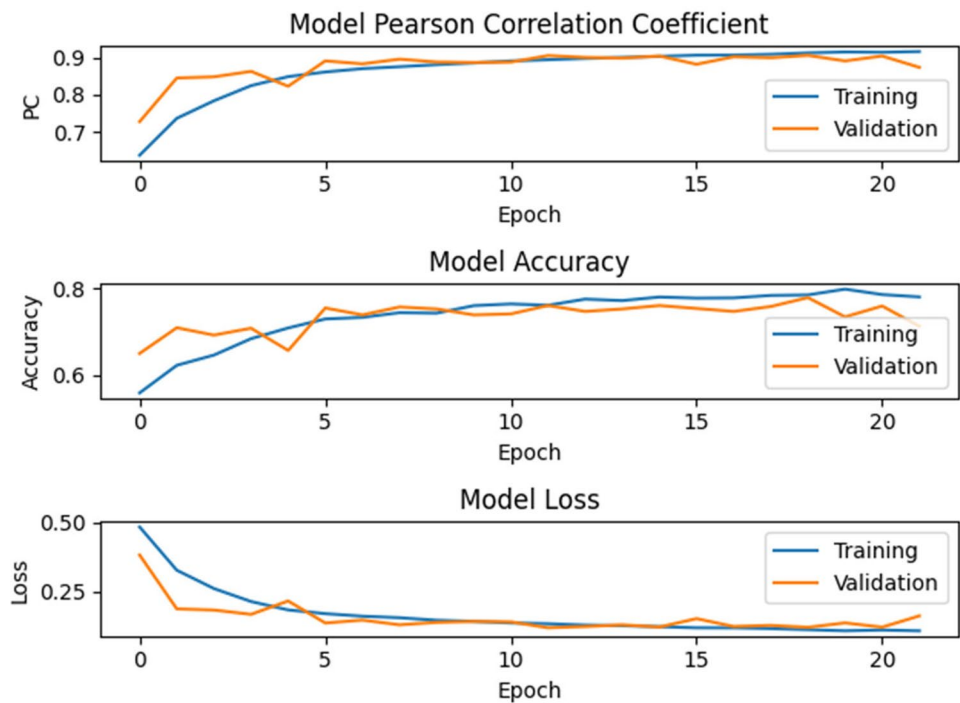
**Fig. 10** CNN network training



## Training and the Evaluation of the CNN-Based Model

Continuous transfer learning was applied in this approach. First, the dataset was used to train the modified ResNet50 model (further CNN model), where layers were modified such that the model learns to calculate the likelihood of each attractiveness estimate. The CNN model produces the same style of outputs as the MLP model. Both networks produce five probabilities, representing the likelihood of receiving each face beauty estimate (1 to 5). However, the MLP and CNN networks' inputs and learning processes differ. CNN model use a picture with $224 \times 224$ resolution as its input (a human face cut from the photo). Such photographs are accompanied by an estimate of each person's attractiveness. As with the MLP model, these estimates and original human images were generated from the SCUT-FBP5500 dataset. Unlike the MLP model, however, the CNN model's training was affected by

data growth, used for diversifying the quantity of data available for learning by utilizing various data transformations.

Training process also employed the following augmentation transformations: twisting, zooming out, zooming in, and rotating. Figure 10 depicts the results of model training. Pearson's correlation coefficient of 0.865 was obtained after training this model. Table 3 summarizes the training results. Adam [46] optimizer was used to train the model. Adam's optimization function was set to the following parameters: learning frequency - 0.001, $\beta_1$ - 0.9, $\beta_2$ - 0.999. *Early stopping technique* was applied to help avoid overfitting.

Figure 11 shows a confusion matrix used to measure the accuracy of the face attractiveness assessment of 1095 faces from the SCUT-FBP5500 dataset. Figure 12 shows the ROC curve. Because the faces analyzed were picked at random, the expert (true) attractiveness rating was evenly dispersed. Most faces were evaluated by experts based on their average

**Table 3** Results of the CNN network's training

| Faces | Number of images for learning | Number of images for testing | Loss | Accuracy | Pearson's correlation coefficient | MAE | RMSE |
|---|---|---|---|---|---|---|---|
| Asian women | 1602 | 398 | 0.204 | 0.688 | 0.844 | 0.083 | 0.125 |
| Asian men | 1601 | 399 | 0.151 | 0.762 | 0.887 | 0.068 | 0.108 |
| Caucasian women | 601 | 149 | 0.181 | 0.664 | 0.854 | 0.077 | 0.118 |
| Caucasian men | 601 | 149 | 0.154 | 0.839 | 0.895 | 0.070 | 0.104 |
| **Total Average** | **4405** | **1095** | **0.175** | **0.732** | **0.865** | **0.075** | **0.116** |

**Fig. 11** CNN network confusion matrix
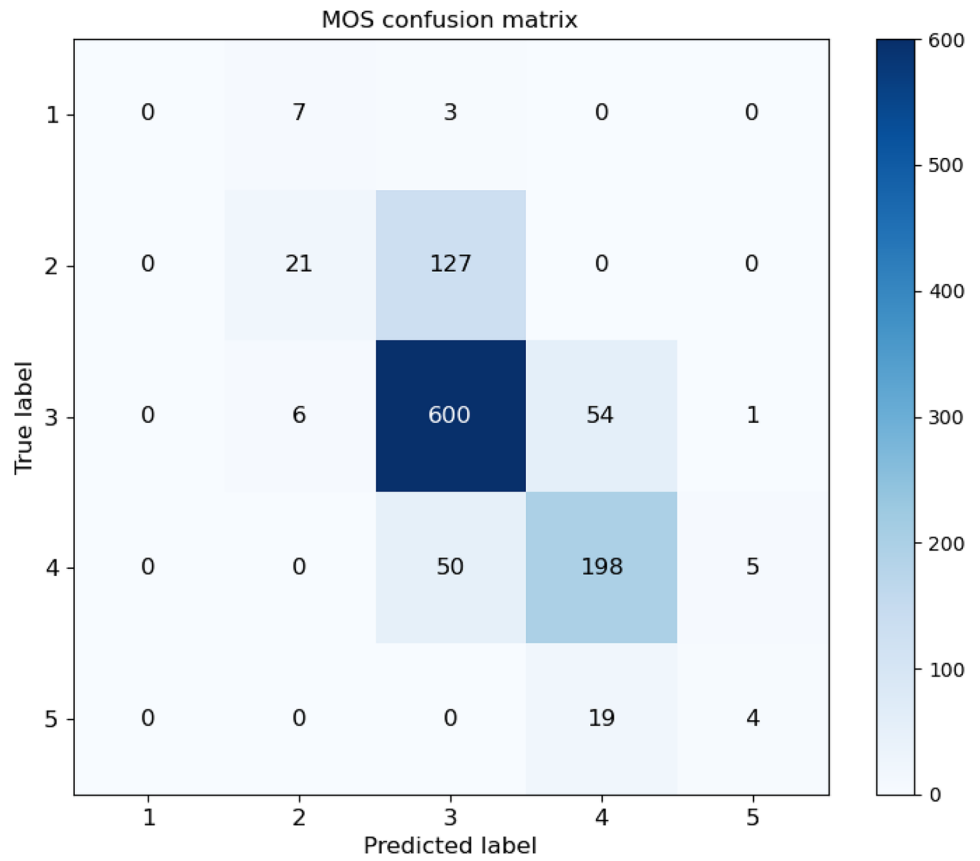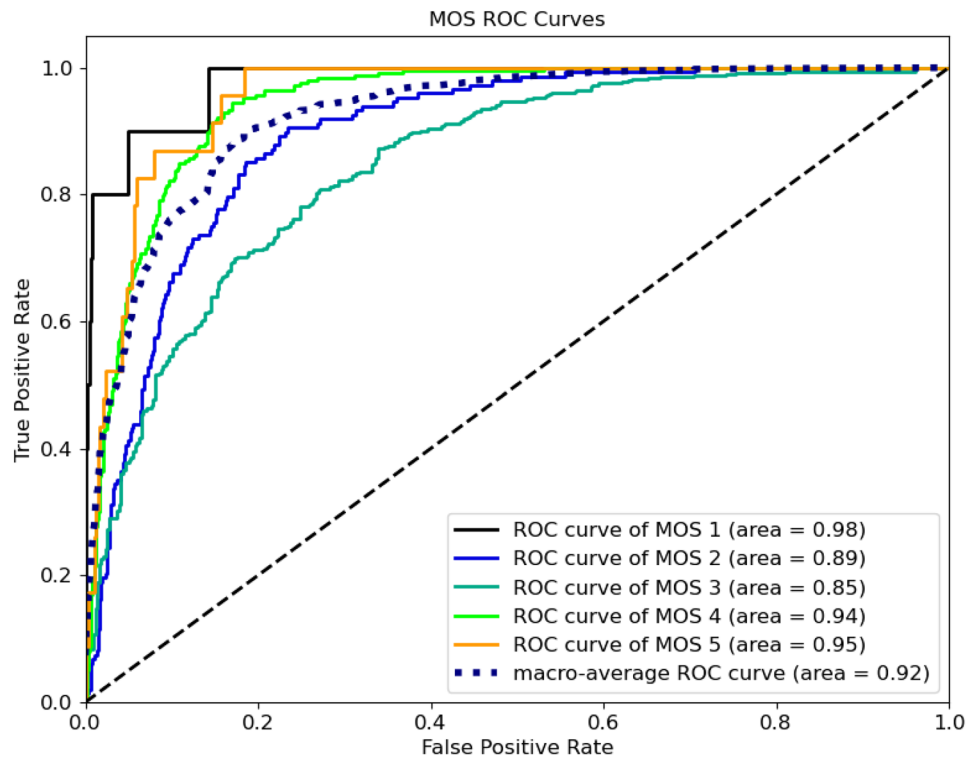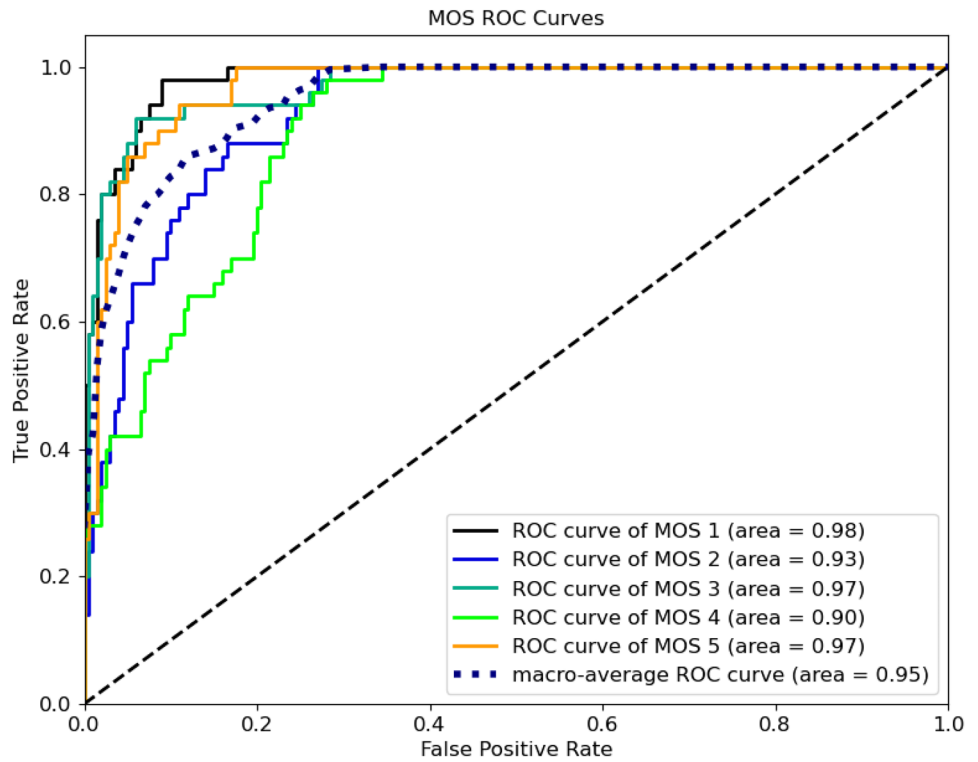


**Fig. 12** CNN network ROC curves

**Fig. 13** CNN network's ROC curves with a sample in which the face attractiveness estimate (MOS) is equally distributed



attractiveness. As a result, most of the faces tested were rated between 2 and 4. The confusion matrix and ROC curves demonstrate the beauty estimate (MOS) determination accuracy for each feasible estimate (1 to 5). According to the confusion matrix, the most common attractiveness score (measured at 600 faces) is 3. This rating is correct, although the CNN network assigns it somewhat more frequently than it should.

The comparison of the accuracy of beauty assessment with a wide distribution of facial attractiveness is presented in Fig. 13, which shows the ROC curves of the network when
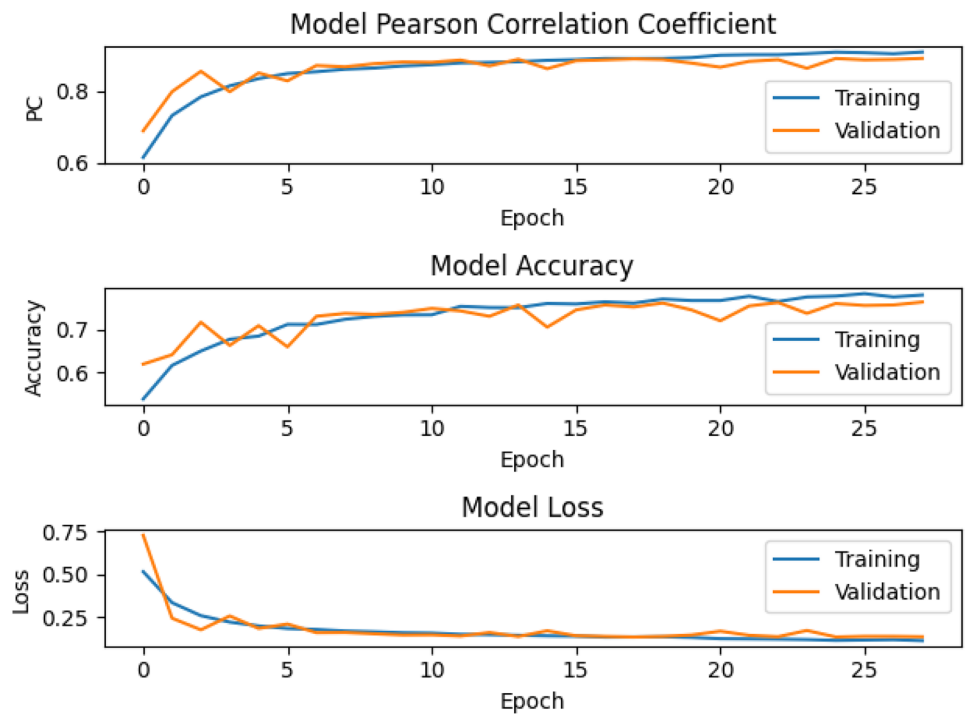
**Fig. 14** CNN network training using generated faces

**Table 4** CNN network training results using generated faces

| Faces | Number of images for testing | Loss | Accuracy | Pearson's correlation | Mae | RMSE |
|---|---|---|---|---|---|---|
| Asian women | 398 | 0.168 | 0.719 | 0.863 | 0.075 | 0.117 |
| Asian men | 399 | 0.137 | 0.777 | 0.896 | 0.066 | 0.104 |
| Caucasian women | 149 | 0.154 | 0.691 | 0.884 | 0.071 | 0.105 |
| Caucasian men | 149 | 0.137 | 0.805 | 0.902 | 0.066 | 0.102 |
| **Total** | **1095** | **0.150** | **0.748** | **0.882** | **0.070** | **0.109** |

250 faces from the SCUT-FBP5500 dataset were used for testing. For each facial beauty assessment, 50 faces were chosen from the data set with attractiveness ratings of 1, 2, 3, 4, and 5. The ROC curve of the CNN network reveals that the area under the average ROC curve improves from 0.76 to 0.95 when compared to the MLP network. The wider the area under the ROC curve, the more precisely the attractiveness of the face is determined.

## Evaluation of the Results Using Generated Faces

After extracting and preprocessing faces from the photos, a "copy" of each face was generated and fed into the CNN network instead of the original photo. The network's training result is shown in Fig. 14. Pearson's correlation coefficient was measured as **0.882.** The results of the training are given in Table 4. Adam optimizer was used to train the network. The Adam optimization function used the following parameters: learning frequency - 0.001, $\beta_1$ - 0.9, $\beta_2$ - 0.999. *Early stopping* was applied to help avoid overtraining.

A confusion matrix is presented in Fig. 15 and ROC plot in Fig. 16. Faces were chosen randomly, to keep the expert (real) beauty estimate naturally distributed. Most faces were evaluated by experts based on their average attractiveness. Most of the faces tested were rated between 2 and 4. The confusion matrix and ROC curves demonstrate the beauty estimate (MOS) determination accuracy for each feasible estimate (1 to 5). The CNN network accurately assigned

**Fig. 15** CNN network confusion matrix using generated faces
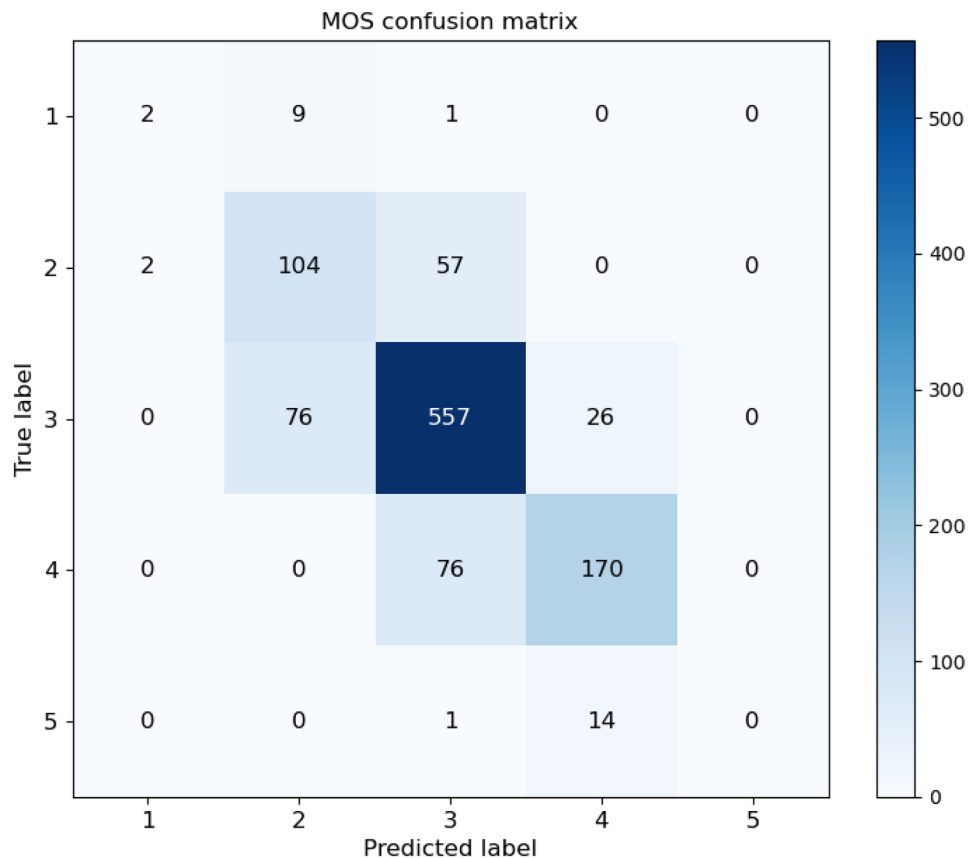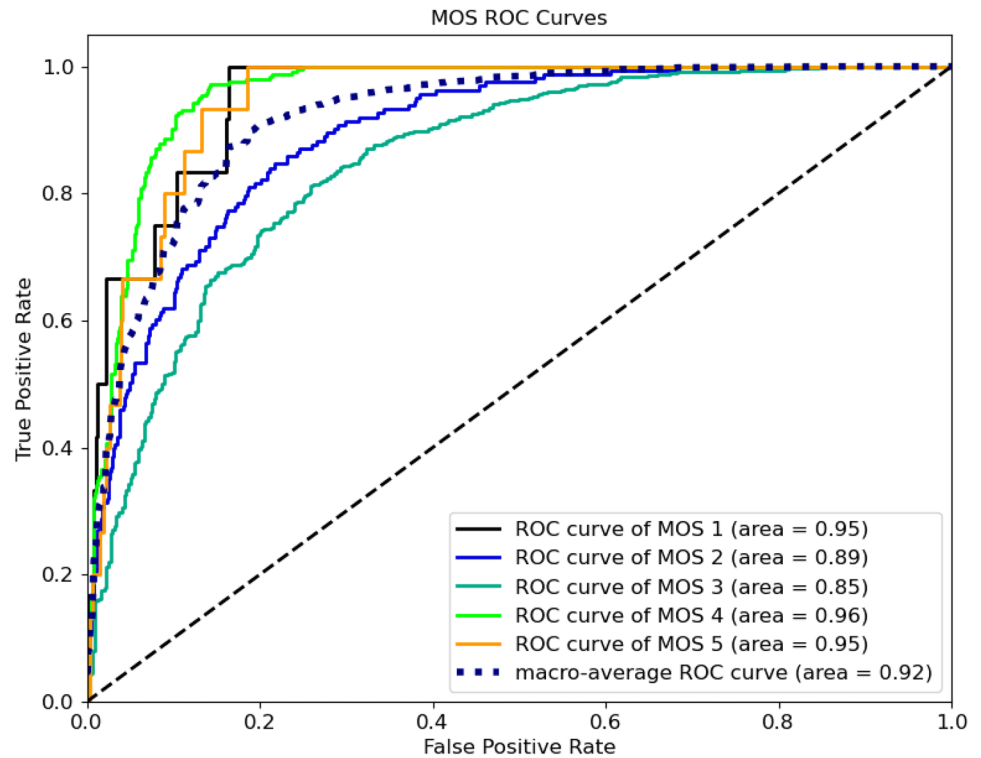
**Fig. 16** CNN network's ROC curves using generated faces



a beauty estimate of 4 to 170 faces, while 40 faces were mistakenly marked in this assessment round. The CNN network also accurately classified the beauty of 104 faces with a beauty score of 2, but 85 faces were estimated with a small mistake. For the least appealing faces with an assessment of one, two faces were correctly recognized and two were incorrectly identified. Even after 15 attempts, the most gorgeous faces could not be identified.

**Fig. 17** CNN network's ROC curves using generated faces with a sample in which the face attractiveness estimate (MOS) is evenly distributed
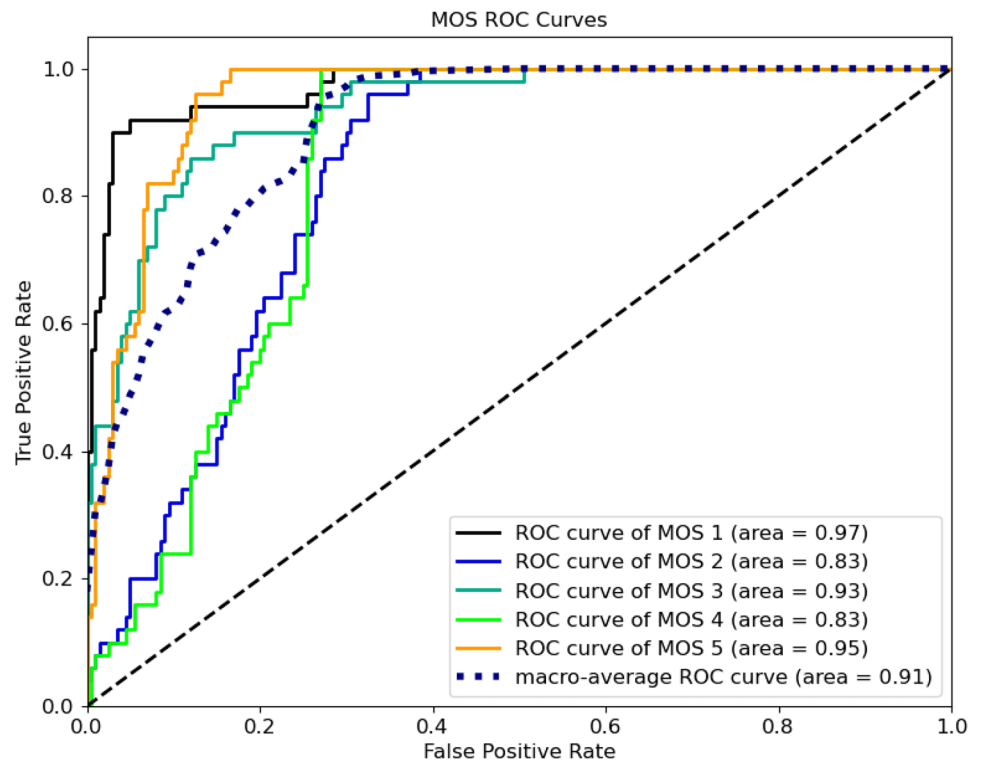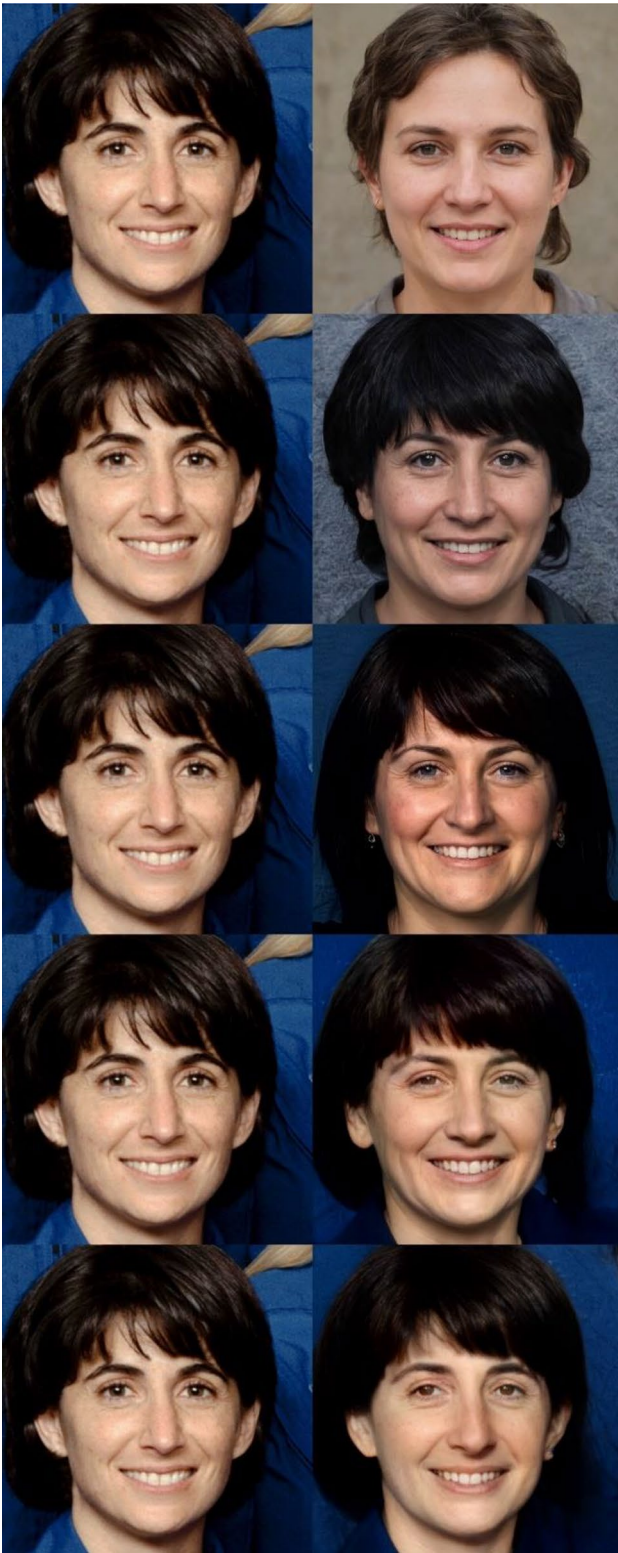
**Fig. 18** Faces generated by GAN in the *latent* space

Figure 17 depicts a comparison of the accuracy of beauty assessment with face attractiveness. It displays the MLP network's ROC curves when 250 faces from the SCUT-FBP5500 dataset were examined. For each facial beauty estimate, 50 faces were chosen from the dataset with attractiveness ratings of 1, 2, 3, 4, and 5. The usage of face synthesis by the CNN network decreases the area under the average ROC curve from 0.95 to 0.91. These findings imply that the CNN network, which employs generated faces, should improve when evaluating faces with beauty ratings of four or two.

## MLP and CNN Network Training to Generate Face Photos

Five thousand five hundred pictures from the SCUT-FBP5500 dataset were added to the latent space to train MLP and CNN networks to judge facial attractiveness by inserting a generated human face. Figure 18 shows an example of a face formed in latent space. An original image is shown on the left side of the figure (always the same). The right side shows how GAN might separate its output in its generation process (so a new face is always different).

Facial attractiveness evaluation without the use of face copies is illustrated in Fig. 19 and facial attractiveness assessment for generated face reproductions is shown in Fig. 20. Both figures show an MLP model beauty estimate ("MLP MOS") and a CNN model beauty estimate ("CNN MOS") for each face.

## Summary of Experiments

We reduced the margin of error in the experiments by employing our CNN-based model. A study of face pictures with varying brightness and contrast, as well as in low resolution, was made to test the robustness. The results have revealed that approach based on the CNN model has less mistakes than the more basic MLP model. This is due to the fact that the MLP model only responds to geometric facial proportions, but the CNN model additionally responds to changes in face color. A summary of the experimental results is given in Table 5. This table shows how different picture parameters impact the overall evaluation. While a change in resolution is a minor signal of a change in evaluation values, a change in brightness, as well as a change in contrast, can have a significant influence on the outcome in many circumstances.
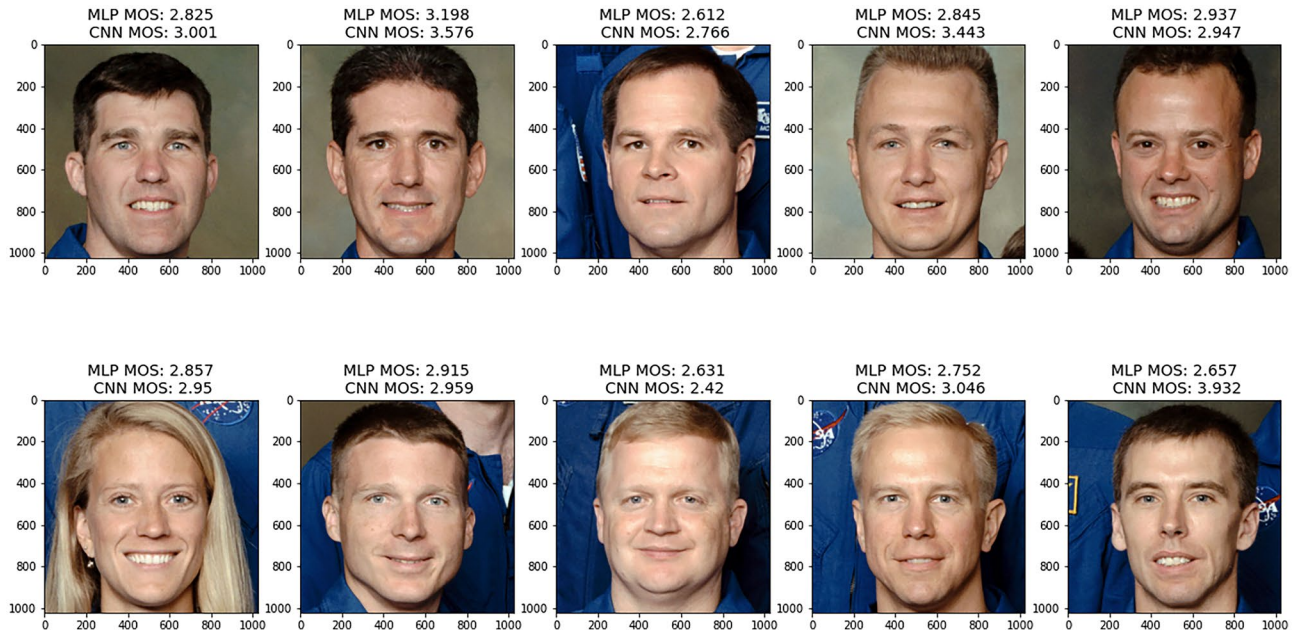
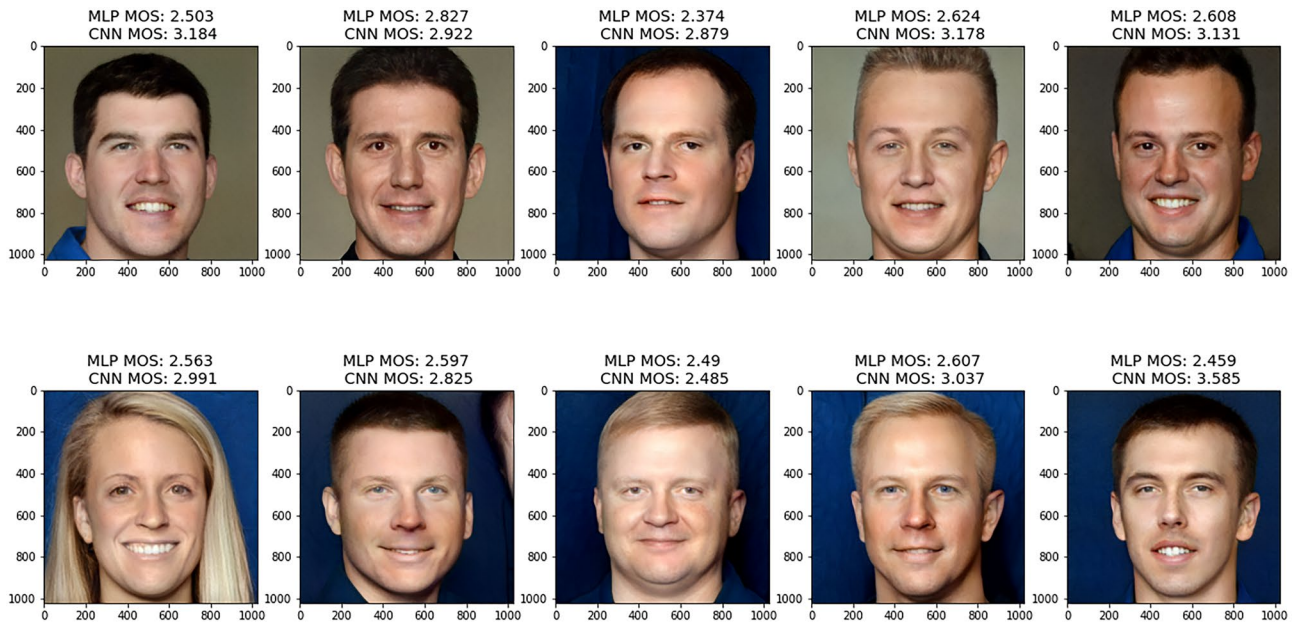**Fig. 19** Measure the attractiveness of faces without generating copies of faces



**Fig. 20** Facial attractiveness rating for generated face copies

**Table 5** Summary of experimental results

| Experiment | MLP MAE | CNN MAE |
|---|---|---|
| Assessment of facial attractiveness in low-resolution photos by cutting the face out of the photo and not applying transformations | 0.483 | 0.247 |
| Assessment of facial attractiveness in low-resolution photos with "super" resolution before cutting off your face | 0.436 | 0.219 |
| Assessment of facial attractiveness in low-resolution photos by cutting the face out of the photo and applying transformations | 0.518 | 0.205 |
| Assessment of facial attractiveness in poor quality photos by cutting the face out of the photo and not applying transformations | 0.437 | 0.316 |
| Assessment of facial attractiveness in poor quality photos with "super" resolution before cutting out the face | 0.440 | 0.493 |
| Assessment of facial attractiveness in poor quality photos by cutting the face out of the photo and applying transformations | 0.453 | 0.318 |
| Assessment of facial attractiveness with different brightness | 0.525 | 0.387 |
| Assessment of facial attractiveness in different contrasts | 0.488 | 0.358 |

## Discussion and Conclusions

Comparison with the state-of-the-art methods using SCUT-FBP5500 dataset is offered in Table 6. The comparison indicates that our methodology outperforms most state-of-the-art algorithms (MAE score of 0.205), with comparable accuracy to regression ensemble–based CNN (MAE score of 0.201), which integrates the ResneXt-50 and Inception-v3 architectures through FC layers but lacks an internal GAN-based evaluator like ours, thus theoretically having a potentially lover internal complexity when approaching unseen faces.

We suggested a novel approach for determining the attractiveness of a face by generating artificial faces. The model's hidden layers can potentially learn useful face traits that are congruent with human visual perception. When it comes to risks to internal authenticity, the network depth is critical. There are two dimensions to external validity threats. On the one hand, training labels are evaluated by certain pupils in a specific community, which may not cover a general perspective in using a simulated face instead of the real face enhances the accuracy of determining facial attractiveness. The training photos, on the other hand, are drawn from Asian and Caucasian population (dependent on the benchmark dataset used), which may lead to a potential data bias in diversity. The capacity to evaluate face beauty also opens the door to hidden facial beauty alterations. The incorporation of the GAN as a component naturally adds some computing cost to the process when compared to pure CNN-based models, which do not use extra or augmented information and rely only on vast, precisely annotated databases for training. However, we feel that refining pruning approaches and minimizing training time is an area that requires attention, as is the adaptation of GAN structures, which might benefit from enhanced sparsity and selectiveness.

Unfortunately, the presented solution may have race-based bias due to composition of a dataset used for training. A more diverse dataset representing more racial types of human faces may be needed to avoid the own-race bias problem [50] and achieve more fair results.

Future study is needed to enhance face normalization in the latent space to increase the accuracy of determining facial attractiveness. Finally, despite improvements shown in recent studies dealing with attractiveness accuracy score due to the non-linearity of deep feature representations, no model is yet sufficiently robust for face beauty evaluation in unconstrained environments, with photos taken from different angles, such as the side, top, or bottom.

## Declarations

**Table 6** MAE comparison with state-of-the-art approaches

| Method | MAE score |
|---|---|
| Label distribution–based CNN [47] | 0.220 |
| ResNet-18 [21] | 0.224 |
| ResneXt-50-R3CNN [48] | 0.212 |
| Original ResNet XT-50 [49] | **0.229** |
| Channel-wise attention CNN [20] | **0.229** |
| *Our GAN-based approach* | *0.205* |
| **Ensemble regression–based CNR** [37] | **0.201** |

## References

1. Little AC, Jones BC, DeBruine LM. Facial attractiveness: evolutionary based research. Philosophical transactions of the Royal

Society of London. Series B, Biol Sci. 2011;366(1571), 1638–1659. https://doi.org/10.1098/rstb.2010.0404.

2. Rasti S, Yazdi M, Masnadi-Shirazi MA. Biologically inspired makeup detection system with application in face recognition. IET Biom. 2018;7:530–5. https://doi.org/10.1049/iet-bmt.2018.5059.

3. Cunningham MR, Barbee AP, Philhower CL. Dimensions of facial physical attractiveness: the intersection of biology and culture. In: Rhodes G, Zebrowitz LA, editors. Facial attractiveness: Evolutionary, cognitive, and social perspectives. Ablex Publishing; 2002. p. 193–238.

4. Siagian C, Itti L. Biologically-inspired face detection: non-brute-force-search approach. Conference on Computer Vision and Pattern Recognition Workshop. 2004;2004:62–62. https://doi.org/10.1109/CVPR.2004.308.

5. Tong S, Liang X, Kumada T, Iwaki S. Putative ratios of facial attractiveness in a deep neural network. In Vision Research 2021;Vol. 178, pp. 86–99. Elsevier BV. https://doi.org/10.1016/j.visres.2020.10.001.

6. Saeed J, Abdulazeez AM. Facial beauty prediction and analysis based on deep convolutional neural network: a review. J Soft Comput Data Min. 2021;2(1):1–12.

7. Burusapat C, Lekdaeng P. What is the most beautiful facial proportion in the 21st century? comparative study among miss universe, miss universe thailand, neoclassical canons, and facial golden ratios. Plast Reconstr Surg - Glob Open. 2019;7(2). https://doi.org/10.1097/GOX.0000000000002044.

8. Rizvi QM, Karawia AA, Kumar S. Female facial beauty analysis for assesment of facial attractivness. Proceedings of the 2013 2nd International Conference on Information Management in the Knowledge Economy, IMKE. 2013;156–160.

9. Carvajal J, Wiliem A, Sanderson C, Lovell B. Towards miss universe automatic prediction: the evening gown competition. Int Conf Pattern Recognit. 2016;1089–1094. https://doi.org/10.1109/ICPR.2016.7899781.

10. Chen F, Xiao X, Zhang D. Data-driven facial beauty analysis: prediction, retrieval and manipulation. IEEE Trans Affect Comput. 2018;9(2):205–16. https://doi.org/10.1109/TAFFC.2016.2599534.

11. Dornaika F, Elorza A, Wang K, Arganda-Carreras I. Image-based face beauty analysis via graph-based semi-supervised learning. Multimedia Tools and Applications. 2020;79(3–4):3005–30. https://doi.org/10.1007/s11042-019-08206-8.

12. El Rhazi M, Zarghili A, Majda A, Bouzalmat A, Oufkir AA. Facial beauty analysis by age and gender. Int J Intell Syst Technol Appl. 2019;18(1–2):179–203. https://doi.org/10.1504/IJISTA.2019.097757.

13. Abayomi-alli OO, Damaševicius R, Maskeliunas R, Misra S. Few-shot learning with a novel voronoi tessellation-based image augmentation method for facial palsy detection. Electronics. 2021;10(8). https://doi.org/10.3390/electronics10080978.

14. Hong Y, Nam GP, Choi H, Cho J, Kim IJ. A novel framework for assessing facial attractiveness based on facial proportions. Symmetry. 2017;9(12). https://doi.org/10.3390/sym9120294.

15. Kaya KS, Türk B, Cankaya M, Seyhun N, Coşkun BU. Assessment of facial analysis measurements by golden proportion. Braz J Otorhinolaryngol. 2019;85(4):494–501. https://doi.org/10.1016/j.bjorl.2018.07.009.

16. Young P. Assessment of ideal dimensions of the ears, nose, and lip in the circles of prominence theory on facial beauty. JAMA Facial Plastic Surgery. 2019;21(3):199–205. https://doi.org/10.1001/jamafacial.2018.1797.

17. Zhang L, Zhang D, Sun M, Chen F. Facial beauty analysis based on geometric feature: toward attractiveness assessment application. Expert Syst Appl. 2017;82:252–65. https://doi.org/10.1016/j.eswa.2017.04.021.

18. Lebedeva I, Guo Y, Ying F. MEBeauty: a multi-ethnic facial beauty dataset in-the-wild. Neural Comput Appl. 2021. https://doi.org/10.1007/s00521-021-06535-0.

19. Wei W, Ho ESL, McCay KD, Damaševičius R, Maskeliūnas R, Esposito A. Assessing facial symmetry and attractiveness using augmented reality. Pattern Anal Appl. 2021. https://doi.org/10.1007/s10044-021-00975-z.

20. Cao K, Choi K, Jung H, Duan L. Deep learning for facial beauty prediction. Information. 2020;11(8). https://doi.org/10.3390/INFO11080391.

21. Lin L, Liang L, Jin L. Regression guided by relative ranking using convolutional neural network (R3CNN) for facial beauty prediction. IEEE Trans Affect Comput. 2019. https://doi.org/10.1109/TAFFC.2019.2933523.

22. Xu J, Jin L, Liang L, Feng Z, Xie D, Mao H. Facial attractiveness prediction using psychologically inspired convolutional neural network (PI-CNN). IEEE Int Conf Acoust Speech and Signal Process. 2017;1657–1661. https://doi.org/10.1109/ICASSP.2017.7952438.

23. Siddiqi MH, Khan K, Khan RU, Alsirhani A. Face image analysis using machine learning: a survey on recent trends and applications. Electronics. 2022;11(8). https://doi.org/10.3390/electronics11081210.

24. Iyer TJ, Rahul K, Nersisson R, Zhuang Z, Joseph Raj AN, Refayee I. Machine learning-based facial beauty prediction and analysis of frontal facial images using facial landmarks and traditional image descriptors. Comput Intell Neurosci. 2021. https://doi.org/10.1155/2021/4423407.

25. Dantcheva A, Dugelay J-L. Assessment of female facial beauty based on anthropometric, non-permanent and acquisition characteristics. Multimed Tools Appl. 2014;74(24):11331–55. https://doi.org/10.1007/s11042-014-2234-5.

26. Packiriswamy V, Kumar P, Rao M. Identification of facial shape by applying golden ratio to the facial measurements: an interracial study in malaysian population. N Am J Med Sci. 2012;4(12):624–9. https://doi.org/10.4103/1947-2714.104312.

27. Little AC, Jones BC, DeBruine LM. Facial attractiveness: evolutionary based research. Philosophical transactions of the Royal Society of London. Series B, Biol Scie. 2011;366(1571), 1638–1659. https://doi.org/10.1098/rstb.2010.0404.

28. Mealey L, Bridgestock R, Townsend G. Symmetry and perceived facial attractiveness. J Pers Soc Psychol. 1999;76:151–8. https://doi.org/10.1037/0022-3514.76.1.151.

29. Ishi H, Jiro G, Kamachi M, Mukaida S, Akamatsu S. Analyses of facial attractiveness on feminised and juvenilised faces. Perception. 2004;33(2):135–45. https://doi.org/10.1068/p3301.

30. Foo Y, Simmons L, Rhodes G. Predictors of facial attractiveness and health in humans. Sci Rep. 2017;7:39731. https://doi.org/10.1038/srep39731.

31. Ibáñez-Berganza M, Amico A, Loreto V. Subjectivity and complexity of facial attractiveness. Sci Rep. 2019;9(1). https://doi.org/10.1038/s41598-019-44655-9.

32. Lin L, Liang L, Jin L, Chen W. Attribute-aware convolutional neural networks for facial beauty prediction. In 28th International Joint Conference on Artificial Intelligence (IJCAI-19). 2019. https://doi.org/10.24963/ijcai.2019/119.

33. Gan J, Xiang L, Zhai Y, Mai C, He G, Zeng J, Bai Z, Donida Labati R, Piuri V, Scotti F. 2M BeautyNet: facial beauty prediction based on multi-task transfer learning. IEEE Access. 2020;8:20245–56. https://doi.org/10.1109/access.2020.2968837.

34. Anderson R, Gema AP, Suharjito, Isa SM. Facial attractiveness classification using deep learning. 2018 Indonesian Association for Pattern Recognition International Conference (INAPR). 2018. https://doi.org/10.1109/inapr.2018.8627004.

35. Vahdati E, Suen CY. Facial beauty prediction from facial parts using multi-task and multi-stream convolutional neural networks. Intern J Pattern Recogniti Artif Intell. 2021;35(12). https://doi.org/10.1142/S0218001421600028.

36. Xiao Q, Wu Y, Wang D, Yang Y, Jin X. Beauty3DFaceNet: deep geometry and texture fusion for 3D facial attractiveness prediction.

Computers and Graphics (Pergamon). 2021;98:11–8. https://doi.org/10.1016/j.cag.2021.04.023.

37. Bougourzi F, Dornaika F, Taleb-Ahmed A. Deep learning based face beauty prediction via dynamic robust losses and ensemble regression. Knowl-Based Syst. 2022;242. https://doi.org/10.1016/j.knosys.2022.108246.

38. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial networks. In International Conference on Neural Information Processing Systems (NIPS 2014). 2014;pp. 2672–2680.

39. Rosado P, Fernández R, Reverter F. GANs and artificial facial expressions in synthetic portraits. Big Data Cogn Comput. 2021;5:63. https://doi.org/10.3390/bdcc5040063.

40. Kelly T, Guerrero P, Steed A, Wonka P, Mitra NJ. FrankenGAN. ACM Trans Graph. 2019;37(1):1. https://doi.org/10.1145/3272127.3275065.

41. Dirvanauskas D, Maskeliūnas R, Raudonis V, Damaševičius R, Scherer R. HEMIGEN: human embryo image generator based on generative adversarial networks. Sensors. 2019;19(16). https://doi.org/10.3390/s19163578.

42. Gary BH, Marwan M, Berg T, Learned-Miller E. Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, Marseille, France. 2008.

43. KING DE. Dlib-ml: a machine learning toolkit. Journal Mach Learn Res. 2009;vol. 10. pp. 1755–1758.

44. Tero K, Miika A, Janne H, Samuli L, Jaakko L, Timo A. Training generative adversarial networks with limited data. In Proceedings of the 34th International Conference on Neural Information Processing Systems (NIPS'20). Curran Associates Inc., Red Hook, NY, USA. 2020;Article 1015, 12104–12114.

45. Siahaan E, Redi JA, Hanjalic A. Beauty is in the scale of the beholder: comparison of methodologies for the subjective assessment of image aesthetic appeal. Sixth International Workshop on Quality of Multimedia Experience (QoMEX). 2014;2014:245–50. https://doi.org/10.1109/QoMEX.2014.6982326.

46. Kingma DP, Ba J. Adam: a method for stochastic optimization. arXiv 2017. arXiv:1412.6980.

47. Fan Y-Y, et al. Label distribution-based facial attractiveness computation by deep residual learning. IEEE Trans Multimedia. 2018;20(8):2196–208. https://doi.org/10.1109/TMM.2017.2780762.

48. Lin L, Liang L, Jin L, Chen W. Attribute-aware convolutional neural networks for facial beauty prediction. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence. 2019. https://doi.org/10.24963/ijcai.2019/119.

49. Liang L, Lin L, Jin L, Xie D, Li M. SCUT-FBP5500: a diverse benchmark dataset for multi-paradigm facial beauty prediction, 2018 24th International Conference on Pattern Recognition (ICPR). 2018;pp. 1598–1603. https://doi.org/10.1109/ICPR.2018.8546038.

50. Wong HK, Stephen ID, Keeble DRT. The own-race bias for face recognition in a multiracial society. Front Psychol 2020; (Vol. 11). https://doi.org/10.3389/fpsyg.2020.00208.