IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 40, No. 1, February, 2023, pp. 55-63
Journal homepage: http://iieta.org/journals/ts

An Ensemble DCNNs-Based Regression Model for Automatic Facial Beauty Prediction and Analyzation

Jwan Najeeb Saeed[1*], Adnan Mohsin Abdulazeez[2], Dheyaa Ahmed Ibrahim[3]

[1] Technical Informatics College of Akre, Duhok Polytechnic University, Duhok 42001, Kurdistan Region, Iraq
[2] Technical College of Engineering-Duhok, Duhok Polytechnic University, Duhok 42001, Kurdistan Region, Iraq
[3] Communications Engineering, Techniques Department, Information Technology College, Imam Ja'afar AlSadiq University, Baghdad 10001, Iraq

Corresponding Author Email: jwan.najeeb@dpu.edu.krd

https://doi.org/10.18280/ts.400105

Received: 13 December 2022
Accepted: 10 February 2023

Keywords:
facial beauty prediction, CNN-based regression, facial image attractiveness analysis, ensemble learning, Grad-CAM, attention mechanism

ABSTRACT

One of the most effective social aspects of the human face is its attractiveness. Automatic facial beauty prediction (FBP) is an emerging research area that has gained much interest recently. However, identifying the significant facial traits and attributes that can contribute to the process of beauty attractiveness estimation is one of the main challenges in this research area. Furthermore, learning the beauty pattern from a relatively small, imbalanced dataset is another concern that needs to be addressed. This research proposes an ensemble-based regression model that integrates judgments made by three various DCNNs, each with a different structure representation. The proposed method efficiently predicts the beauty score by leveraging the strengths of each network as a complementary data source, and it draws attention to the most important beauty-related face features through the Gradient-weighted Class Activation Mapping (Grad-CAM). The findings are promising, demonstrating the efficiency of fusing the decision of multiple predictors of the proposed ensemble DCNNs regression models that is significantly consistent with the ground truth of the employed datasets (SCUT-FBP, SCUT-FBP5500, and ME Beauty). Moreover, it can assist in comprehending the relationship between facial characteristics and the impression of attractiveness.

1. INTRODUCTION

Recently, beauty-related sectors have grown dramatically around the world. Researchers from various fields, including the entertainment industry, digital media, plastic surgery, the cosmetics industry, and artists, have conducted several studies to analyze and quantify face attractiveness [1]. Predicting facial attractiveness is a significant and complicated task in computer vision and machine learning. Furthermore, building a robust and effective face beauty estimation model is difficult because of the variability of facial appearance and the complexity of human perception. FBP has two main folds issues that need to be addressed. The first fold is identifying the accurate face representation, which can be either feature-based, holistic, or hybrid feature representation. The second fold is FBP model might be easily prone to over-fitting due to the lack of adequate and efficient face datasets.

Geometric facial characteristics, color, texture, and other local representations are all examples of feature-based representation. In addition, geometric features take into account the locations of facial landmarks, the distances between them, or the relationship between these distances [2]. Early FBP studies have used hand-crafted feature descriptors that could be utilized for facial appearance. This is because a human's judgment of facial beauty is influenced by facial appearance factors such as skin texture and color and gender, race, and age. On the other hand, holistic techniques emphasize global information about the face rather than local aspects [3]. However, utilizing statistical and conventional machine learning techniques for beauty feature extraction and prediction has become less effective with advanced deep learning neural networks (DNNs).

Deep learning is extremely useful in computer vision applications [4, 5]. Researchers have found that deep data-driven approaches such as convolutional neural networks (CNNs) have proved their efficiency in computing the attractiveness of facial images either as an extractor or predictor. The CNN model with a deeper structure, larger input images, and smaller convolution kernels achieves better performance. However, deep learning needs a considerable amount of data that is often scarce. Therefore, transfer learning can minimize the reliance on massive volumes of data and mitigate overfitting issues. Meanwhile, training the network from scratch can sometimes produce slightly better results than fine-tuning a network. This suggests that, if computation is not a restriction, the model trained from scratch outperforms the fine-tuned ones in certain scenarios and configuration settings [6]. The reason behind this is that the well-known pre-trained networks have been trained on other related but different tasks, such as object identification, image classification, etc. Meanwhile, training a network from scratch on a dedicated task can provide more suitable weights for the learning process of the target task.

In deep and machine learning contexts, CNNs are viewed as "black-box" processes, along with many other techniques, where users can analyze and interpret both the input and the output, but the algorithm's explanation is vague. This may lower the user trust in the outcomes and influence decision-making. The primary goal of developing facial feature analytical techniques was to create a link between feature representations needed by real-world computer vision tasks and visual explanations that are understandable by humans. Therefore, Grad-CAM technique is utilized to evaluate the most discriminative features discovered by CNN.

FBP could be a classification, regression, or ranking problem. Both regression and classification models are crucial in predictive analytics, particularly machine learning, and artificial intelligence [7]. Facial beauty assessment is viewed as learning continuous targets (beauty score) from a restricted dataset with an imbalanced label distribution. To attain the objective, a DCNN-based regression is required to be constructed.

The process of producing many models of a regression network and combining them to generate an ensemble prediction is known as ensemble learning. This method is useful for enhancing the overall performance of individual networks [8, 9]. This work utilizes different learning environments through the concept of pre_trained network's knowledge transferring and training a new network from scratch and then ensembling them to present the following key contributions:

- Proposing an ensemble DCNNs-based regression model to learn the continuous value (beauty score) of the facial image. The suggested model consists of one network trained from scratch and two fine-tuned well-known pre-trained CNNs, namely AlexNet and VGG16.
- Utilizing the decision fusion process based on the average of these sub-models. It makes a judge based on the consensus of three different CNNs architectures and various learning environments to automatically predict the facial beauty score of both genders with diverse ages, poses, and ethnicities.
- Unifying the beauty scores range to be (1-5) for all utilized datasets. Consolidating the beauty score range can assist in comprehending and analyzing model efficiency, especially when adopting multiple datasets with various beauty score ranges.
- Employing attention mechanism and feature visualization to identify the beauty-related features via Grad-CAM to evaluate the most discriminative features discovered by the proposed model.

Three dedicated FBP datasets (SCUT-FBP, SCUT-FBP 5500, MEBeauty) are used to show the effectiveness of the proposed model. The results are promising because they show the effectiveness of combining the evaluations of different predictors in the proposed ensemble DCNNs regression model, which is significantly consistent with the ground truth of the datasets used. It also can analyze the relationship between facial features and the perspective of attractiveness.

The rest of this paper is structured as follows. Section 2 provides a quick overview of the relevant work. Section 3 discusses the proposed model's methodology. Section 4 contains a discussion of the obtained results. Finally, section 5 shows the conclusion.

2. RELATED WORK

Early investigations on FBP were mostly concerned with feature engineering and based on hand-rafted features (e.g., putative ratios and geometric features [10, 11], apparent features [12]). Then the traditional machine learning-based classifier or regressor (e.g., support vector machine [13], artificial neural network [14], and the k-nearest neighbors [15]) were used to fit the hand-crafted features. However, before the advent of deep learning, feature engineering was necessary for computer vision applications. With the rapid growth of intelligent applications today, there is a growing demand for the automatic extraction of biometric information from facial images. Using CNN, an end-to-end learning technique can be created by learning the mapping from the input to the desired output.

In recent years, many researchers have focused on CNNs as a new machine learning research technique [16]. Deep convolutional neural networks (DCNNs) outperform hand-crafted descriptors in terms of feature extraction and prediction. Some CNN framework-based models, such as VGG [17, 18], ResNet [19], were used to represent the hand-crafted features. A multi-task CNN called HMTNet was proposed in the study [20] that can predict the beauty score of a facial image besides race and gender utilizing SCUT-FBP and SCUT-FBP5500 datasets. Furthermore, a new dataset was proposed in the study [21], and the knowledge of DenseNet, Xception pre-trained on ImageNet, and VGG16 that pre-trained on VGGFace2 dataset was transferred to find the best model for FBP. To make deep neural networks deployable on compact hardware, Saeed et al. [22] built a light-deep CNN for assessing the attractiveness of facial images from scratch called FIAC-Net and evaluated the presented network on SUT-FBP, SUT-FBP5500, and CelebA datasets.

3. THE PROPOSED MODEL

Machine-based facial image beauty assessment like a human prediction, is a new and challenging research field. CNN has made considerable progress in computer vision in recent years due to its incredible capacity to learn discriminative features. Generally, CNNs consist of several convolutional layers, pooling layers, nonlinear activation functions, and ultimately one or more fully connected layers.

While deep learning requires a significant amount of samples, datasets within FBP are relatively small. Therefore, reducing the dependency on large data sets and overfitting issues necessitates transferring the knowledge of previously trained networks on distinct but related tasks to be fine-tuned with the new mission of predicting the score of the facial image aesthetic. On the other hand, sometimes starting from scratch is better than fine-tuning a network. If computation is not a limitation, the model trained from scratch outperforms the fine-tuned ones under certain scenarios and configuration settings. To combine the merits of both aforementioned deep learning approaches, this work proposed an ensemble regression deep convolutional neural network model to automatically assess the facial image aesthetic score. In addition, the proposed model makes a decision based on the average of three CNNs that were learned utilizing various datasets and learning environments, as depicted in the next subsections.

3.1 Data augmentation

FBP datasets show that data on facial image beauty is frequently skewed. This is because most people have an average level of beauty. This type of imbalance is known as intrinsic imbalance and is caused directly by the nature of the data space. For instance, the most rated beauty scores in both SCUT-FBP and SCUT-FBP5500 datasets ranged [2-2.9] out of 5 scores. Similarly, the dominant beauty score in MEBeauty dataset ranges [4.0-4.9] out of 10 scores which is equivalent to the beauty scores averages of the two earlier datasets. The remaining beauty score rates are relatively less frequent than those aforementioned ranges.

The small sample size and imbalanced data together provide a new challenge for estimating facial image aesthetics. In this work, both online and offline data augmentation techniques are employed to tackle the issue of small imbalanced datasets. While online augmentation is done through model training time, offline augmentation is conducted by transforming the training set images in the pre-processing phase. Various augmentation strategies have been considered in this study. For instance, standard reflections as rotation, a more extensive set of augmentation as color transformation, and synthetic noise that improves the tolerance for input quality variations caused by the addition of noise, as shown in Figure 1.

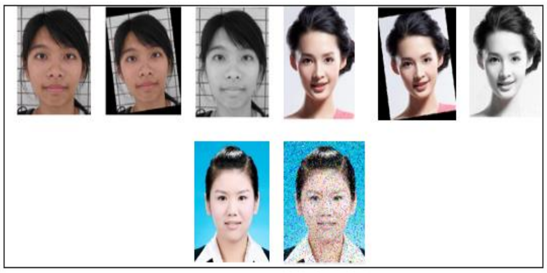


Figure 1. Samples of SCUT_FBP data augmentation

FBP datasets may have different ranges of beauty scores. For instance, both SCUT-FBP and SCUT-FBP5500 have beauty scores ranging [1-5] per image. While ME Beauty dataset has a beauty score ranging [1-10]. Consequently, Unifying the range of the beauty score can help to understand and analyze the model efficiency, particularly when implementing more than one dataset with different beauty score ranges. The proposed method unifies the beauty score to be [1-5] for the three used datasets.

For Beauty-Reg-Net, the input image is resized to 128×128. Furthermore, the image input size of the VGG16 is 224×224. While AlexNet has 227×227 pixels for the input layer. Then the image is passed to each network individually. Figure 2 illustrates the framework of the proposed model, which includes the following aspects:

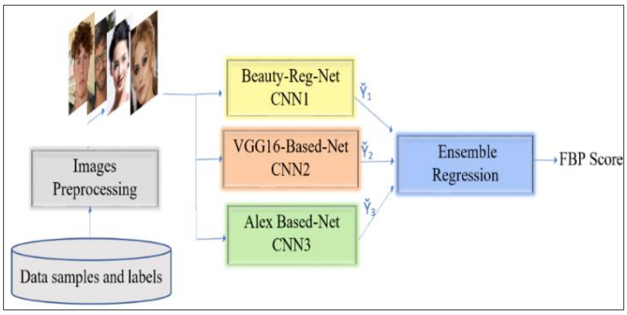


Figure 2. The general framework of the proposed model

3.2 Beauty-reg-net

It is a light DCNN built and trained from scratch with seven trainable layers, six convolutional layers for the feature maps creation, and one fully connected layer to predict the facial image beauty score. Typically, the activation functions followed convolutional layers. The Beauty-Reg-Net first applies batch normalization (BN) to ensure the regularization and decrease the generalization error, followed by the rectified linear unit (ReLU) activation function for all layers except the final layer that used mean-secured-error with the response as a loss function. In addition, the feature map of the preceding layer is downscaled using pooling layers. It keeps track of where the kernels and good matches are.

With max-pooling, the highest value in a patch is added to the new image. When employing average pooling, the average of all values is taken and put in the corresponding location of the output matrix. The proposed Beauty-Reg-Net employs max pooling after each activation function, except for the final ReLU layer, which operates average pooling. Moreover, the dropout strategy is utilized before the fully connected layer (FC) to achieve regularization by dropping out a percentage of 20% from the preceding layer to prevent overfitting and enhance the generalization. For each input face, the network gives a score. The correlation between the actual and the predicted scores is used in evaluating the model performance. Figure 3 depicts the architecture of Beauty-Reg-Net, and the configurations of the main layers are summarized in Table 1.

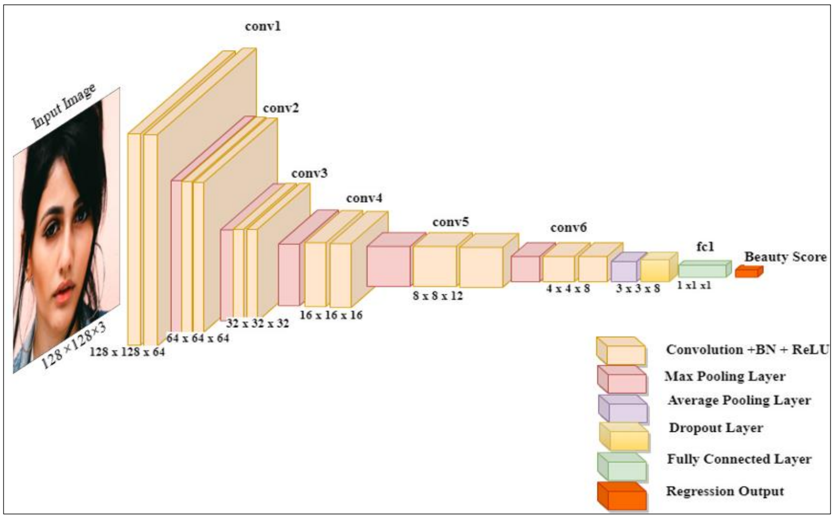


Figure 3. The architecture of the proposed Beauty-Reg-Net

Table 1. The configurations of proposed light-deep Beauty-Reg-Net layers

Layer Name	Kernels	Size	Stride
Input	-	128 × 128 ×3	-
Convolutional-1 (BN+ReLU)	64	11 × 11	1
Max pooling		2 × 2	2
Convolutional-2 (BN+ReLU)	64	9 × 9	1
Max pooling		2 × 2	2
Convolutional-3 (BN+ReLU)	32	7 × 7	1
Max pooling		2 × 2	2
Convolutional-4 (BN+ReLU)	16	5× 5	1
Max pooling		2 × 2	2
Convolutional-5 (BN+ReLU)	12	3 × 3	1
Max pooling		2 × 2	2
Convolutional-6 (BN+ReLU)	8	3 × 3	1
Average pooling		2 × 2	1
Fully Connected + Dropout			
Mean-Secured-Error with response			
Regression			

3.3 Transfer learning and network fine-tuning

It is a technique for transferring knowledge from one or more source tasks to a target task. Transfer learning is the most recent benchmark in the deep learning approach, it is also thought to be able to improve current structures so that they are suitable for new tasks. Furthermore, transfer learning is utilized to overcome over-fitting when the number of training samples is restricted [23]. Moreover, it helps to save training time. The proposed model transfers the knowledge of the networks pre-trained on the ImageNet dataset, which is a massive visual database designed for use in visual object identification software. It is recognized as a significant achievement in machine learning model evaluation. ImageNet has manually annotated over 14 million images and 1000 classes to identify items in at least one million images [24]. The proposed model fine-tunes these networks to be adapted to the facial image beauty estimation task.

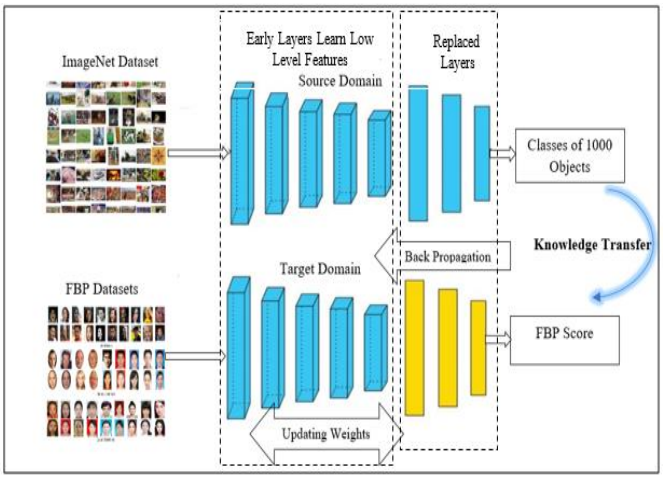


Figure 4. The general notion of transferring the knowledge of the pre-trained CNN

When CNN is fine-tuned, it gets retrained on new data for a distinct but related task by removing the final layer(s) of the network that was previously trained. Then, configure the new layers of the same deleted class. The last fully connected layer (or layers)) in the learnable neural nodes of the classification model is replaced by a new layer with a number of neurons that are comparable to the output of the new task, which is, in our case, one neuron for the regression output to estimate the face aesthetic level. Figure 4 depicts the general framework of the transfer learning process.

3.3.1 VGG16-Net based regression

It is an abbreviation for the Visual Geometry Group proposed in the study [25], a CNN with 16 layers of depth. The construction of VGG networks is simple and frequently utilized in various computer vision applications. It comprises a stack of 2x2 max pooling layers to downscale feature maps by a factor of two after the 3x3 convolution layers for extracting image features. The structure is repeated until the output has a small spatial size. Figure 5 illustrates the architecture of the proposed fine-tuned VGG16-based predictor. The thousand fully connected layers (fc8) were replaced by one fully connected layer (fc). Similarly, the softmax and classification layers were replaced by the MSE with the response and the regression layers to be adapted to the FBP process. The layout of the proposed fine-tuned VGG16-Net is presented in Table 2.

Table 2. The configurations of the proposed fine-tuned VGG16-Net

Layer Name	Kernels	Size	Stride
Input		224×224 ×3	
Convolutional1_1 + ReLU1_2	64	3 × 3	1
Convolutional1_2 + ReLU1_2	64	3 × 3	1
Max pooling_1		2 × 2	1
Convolutional2_1 +ReLU2_1	128	3 × 3	1
Convolutional2_2 +ReLU2_2	128		
Max pooling_2		2 × 2	1
Convolutional3_1 + ReLU3_1	256	3 × 3	1
Convolutional3_2 + ReLU3_2	256	3 × 3	1
Convolutional3_3 + ReLU3_3	256	3 × 3	1
Max pooling_3		2 × 2	1
Convolutional4_1 + ReLU4_1	512	3 × 3	1
Convolutional4_2 + ReLU4_2	512	3 × 3	1
Convolutional4_3 + ReLU4_3	512	3 × 3	1
Max pooling_4		2 × 2	1
Convolutional5_1 + ReLU5_1	512	3 × 3	1
Convolutional5_2 + ReLU5_2	512	3 × 3	1
Convolutional5_3 + ReLU5_3	512	3 × 3	1
Max pooling_5		2 × 2	1
Fully Connected fc6 +ReLU+ Dropout			
Fully Connected fc7+ ReLU+ Dropout			
Fully Connected fc8			
MSE with response			
Regression			

3.3.2 Alex-Net based regressor

It is a CNN proposed by Krizhevsky et al. [26]. It has five convolutional layers, three fully connected layers, two dropout layers, and max-pooling layers for 62.3M learnable parameters. Furthermore, with the exception of the final output layer, which uses the Softmax activation function, the ReLU is the activation function of all layers (see Figure 6). To adapt AlexNet to predict the facial beauty score Softmax is replaced by the mean absolute error that converts the classification process to a regression task. Table 3 shows the configurations of the proposed fine-tuned Alex-Net.

Table 3. The configurations of the proposed fine-tuned Alex-Net

Layer Name	Kernels	Size	Stride
Input		227 × 227 ×3	
Convolutional_1 + BN+ReLU	96	11 × 11	4
Max pooling_1		2 × 2	2
Convolutional_2 + BN+ReLU	256	5 × 5	1
Max pooling_2		2 × 2	2
Convolutional_3 +ReLU	384	3 × 3	1
Convolutional_4+ ReLU	384	3× 3	1
Convolutional_5 +ReLU	256	3 × 3	1
Max pooling_3		2 × 2	2
Fully Connected fc6			
Fully Connected fc7			
Fully Connected fc8			
Mean-Secured-Error with response			
Regression			

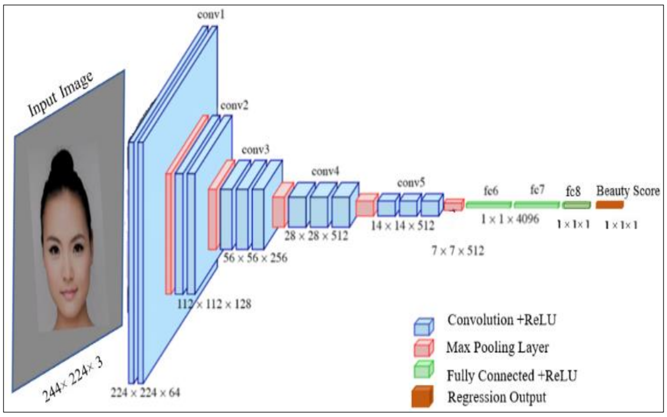


Figure 5. The architecture of the proposed fine-tuned VGG16-Net

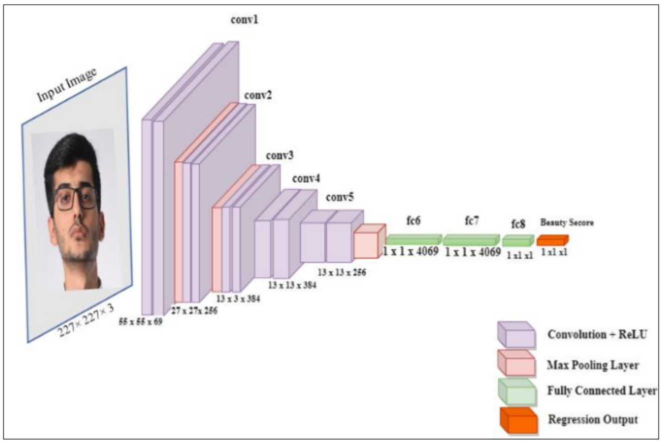


Figure 6. The architecture of the proposed fine-tuned Alex-Net

3.4 Model training

The implementation workflow of the proposed model was started by splitting the data samples into 80% for the training phase and the rest of 20% for the testing phase. During the training stage, an online augmentation is applied to improve the learning process by modifying mini-batches that are fed to the model. This technique can help to reduce an exponential rise in storage needs caused by offline augmentation, which demands storage space. Consequently, training data is rotated up to Twenty degrees and randomly translated in horizontal and vertical dimensions up to three pixels. In addition, a random reflection is performed in the top-bottom and left-right directions. The adaptive movement estimation method (Adam) is implemented as an optimizer due to its effective computations, low memory needs, and capacity to handle sparse gradients on noisy data. The hyper-parameters setting values are shown in Table 4.

Table 4. Training hyper-parameters setting

Hyper-parameter	Setting
InitialLearnRate	0.001
Optimizer	Adam
MaxEpochs	150
MiniBatchSize	64
Shuffle	Every epoch
Verbose	False

3.5 Decision fusion

The ensemble-based approach is developed when various sub-models are trained, and their outputs are combined using predetermined techniques. It has been revealed that this can result in higher performance compared to a single learning model [27]. In order to reach a consensus that is superior to the individual decisions of the predictors, decision fusion seeks to combine the decisions made by various sub-models. It is conducted to improve the performance of the prediction. The decisions of the proposed DCNNs are ensembled. The ultimate score is determined by taking the average of the facial beauty scores that the various sub-models gave. This can contribute to inheriting the merits of each deep network and enhance the performance of the proposed model.

3.6 Gradient-weighted class activation mapping

Since CNNs produce tough results to explain, neural networks (NNs) are often referred to as black boxes. As a result, when NNs provide strange findings, there may be no way to figure out why. Gradient-weighted class activation mapping (Grad-CAM) [28]. It is a way to construct visual explanations for decisions made by CNN-based models to make them more transparent. It generates a heat map that is deployed to a CNN after the training phase, and the parameters are set. Grad-CAM allows us to verify where our network is focusing visually, concentrate on the relevant patterns in the image, and figure out which image features are the most crucial for the classification or prediction process [29, 30]. Using the Grad-CAM technique, the proposed model can provide a relationship between human-understandable visual explanations and feature representations and evaluate the most discriminating features discovered by the ensemble DCNN.

4. RESULTS AND DISCUSSION

An ensemble DCNNs_based regression is implemented for automatic facial image beauty estimation. Three FBP benchmarks with 20% of the test data are utilized to demonstrate the effectiveness of the proposed model.

4.1 Dataset

Three FBP datasets are used to evaluate the effectiveness of the proposed model on male and female facial images with various ages, ethnicities, and constraints, as described below:

- SCUT-FBP [31] is designed for the automatic recognition of facial attractiveness. It contains 500 various good-resolution, frontal photos of Asian women faces with simple backgrounds and neutral expressions. The scores for attractiveness, which range from [1-5], are determined by averaging the ratings of 70 raters for every image.
- SCUT-FBP 5500 [32] is an improved version of SCUT-FBP, and it is used to assess the effectiveness of various facial aesthetic prediction techniques. It includes 5500 frontal faces of different ethnicities (Asian/Caucasian) with various ages of both male and female subjects. Each facial image has an attractiveness score ranging [1-5] that was obtained by taking the average of the ratings given by 60 different raters.
- MEBeauty [21] is a multi-ethnic of 2550 facial images in-the-wild dataset that was created more recently. It includes 1250 male and 1300 female faces of various ages. Black, Asian, Caucasian, Hispanic, Indian, and Middle Eastern faces are included. This diversity can help to eliminate potential racial, ethnic, cultural, or social biases in attractiveness perception. Over 300 volunteers of various ages and racial backgrounds rated the beauty score of these images to be ranged [1-10]. It is expected to be one of the most used datasets for future research on facial aesthetic assessment.

4.2 Metrics of evaluation

The evaluation of FBP performance is measured using three commonly used metrics: Pearson Correlation (PC), which has values ranging from 1 to -1, with 1 indicating total positive linear correlation, 0 showing no linear correlation, and -1 indicating whole negative linear correlation. The mean absolute error (MAE) and the root mean square error (RMSE) are metrics that indicate the efficiency of the model; values toward zero indicate good performance [21, 33]. These measures are defined as follows:

Given a set of N tests in samples:

$$PC=\frac{\sum_{i=1}^{N}(y_i-\bar{y})(p_i-\bar{p})}{\sqrt{\sum_{i=1}^{N}(y_i-\bar{y})^2}\sqrt{\sum_{i=1}^{N}(p_i-\bar{p})^2}} \qquad (1)$$

$$MAE=\frac{1}{N}\sum_{i=1}^{N}|y_i-p_i| \qquad (2)$$

$$\text{RMSE}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-p_i)^2} \qquad (3)$$

The ground truth label is expressed by y_i, and p_i represents the estimated score of the Ith image. Meanwhile, $\bar{y}$ represents the average of all labels in the ground truth and $\bar{p}$ refers to the average of the predicted scores. Generally, higher PC values with lower MAE and RMSE values reflect better performance.

4.3 Experimental setup

The R2022a version of MATLAB was used to run the suggested model on a machine with an Intel(R) Core(TM) i7-10750H CPU and 16G of RAM. The data samples are separated into 80% for training and 20% for testing. Three aforementioned dedicated facial aesthetic benchmarks are used to assess the performance of our proposed ensemble DCNNs regression-based model.

4.4 Performance evaluation

The proposed model tries to utilize a system with multiple predictors to achieve the objective of the most efficient predictors. The intuition behind combining predictors is to improve the model performance instead of implementing each one individually because decision fusion strategies are based on the notion that various estimators will not make the same errors. The comparison of the Pearson correlation coefficients obtained by the proposed ensemble model on the different FPB datasets is shown in Figure 7.

It is obvious from Figure 7 that applying the proposed method to SCUT-FBP gives the PC value of 0.879. Meanwhile, both SCUT-FBP 5500 and MEBeauty datasets show slightly higher PC values at about 0.886 and 0.888, respectively. Moreover, the performance evaluation of the proposed model based on the error rate is shown in Figure 8.

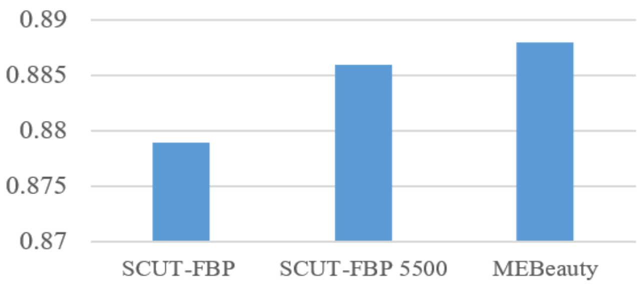


Figure 7. The Pearson correlation comparison of the proposed ensemble model on various FPB datasets

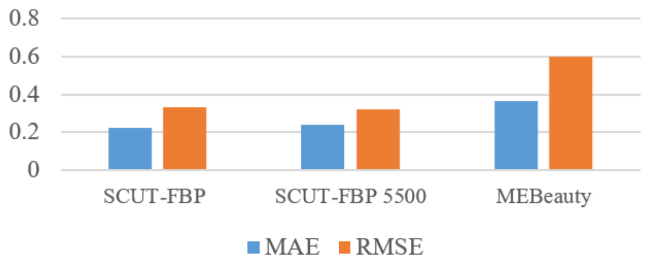


Figure 8. The performance evaluation of the proposed model based on the error rate

The MAE and RMSE of 0.226 and 0.33, respectively, were achieved when utilizing SCUT-FBP, which illustrates good performance compared to the two larger used datasets. Since SCUT-FBP contains frontal Asian female facial images, finding beauty in female facial images with one ethnicity is easier than in datasets with multi-ethnics and both genders. Consequently, the MAE and RMSE that resulted from applying the testing data samples of SCUT-FBP 5500 were 0.242 and 0.320, respectively.

Contrarily, samples in the MEBeauty dataset include wild images and unconstrained conditions with various poses, genders, ages, and races. Thus, it shows the MAE as 0.365 and RMSE as 0.6, which refers to the highest error values among the three datasets.

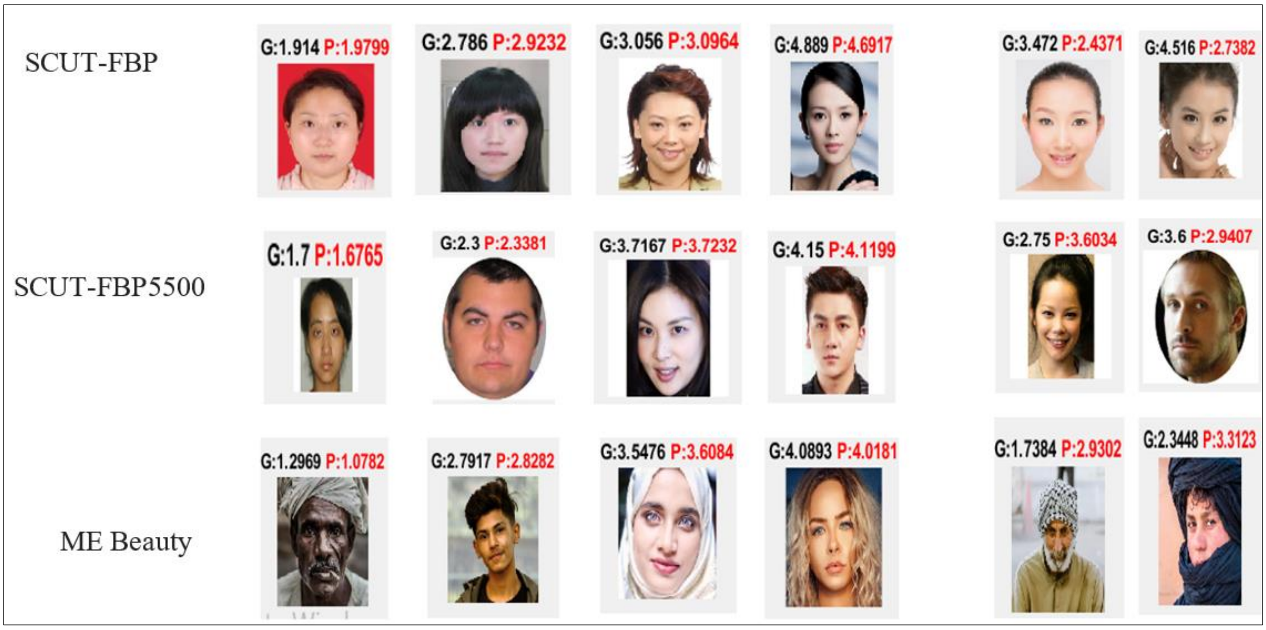
The results show the effectiveness of fusing the assessments of various estimators of the proposed ensemble DCNNs regression model. Figure 9-(a) elucidates the predicted beauty scores of some tested data samples that are significantly comparable to the actual attractiveness rating of the corresponding datasets. However, some aspects, such as non-frontal poses, unneutral expressions, or wearing accessories, may influence the improperly judged samples (see Figure 9-(b)).

4.5 Facial characteristics and attractiveness perception

To understand the relationship between face characteristics and attractiveness perception, we visualized the features of the hidden layers by employing the Grad-CAM technique. Then, it illustrates that face traits such as hairstyle, eyes, and mouth are more beauty-informative and discriminative in most of the tested data (see Figure 10). Most studies on face attractiveness computation have focused on two-dimensional(2D) facial images, specifically the two-dimensional frontal aspect of faces due to the lack of a 3D FBP dataset. However, the heights of the cheekbones and the nose can significantly assist in determining physical beauty. With a 2D frontal facial image, such information is difficult to acquire.

4.6 FBP comparison with state-of-the-art models

To evaluate the efficiency of the proposed model for FBP compared to the existing deep learning models, we compared the performance of our model to that of the previous findings of other studies. The experimental findings in this study show that the proposed ensemble model performed better than the earlier state-of-the-art, as demonstrated in Table 5.



(a) Accurately predicted instances (b) Inaccurately predicted instances

Figure 9. Samples of the predicted tested data based on the proposed model, G: ground truth beauty score; P: Machine predicted beauty score

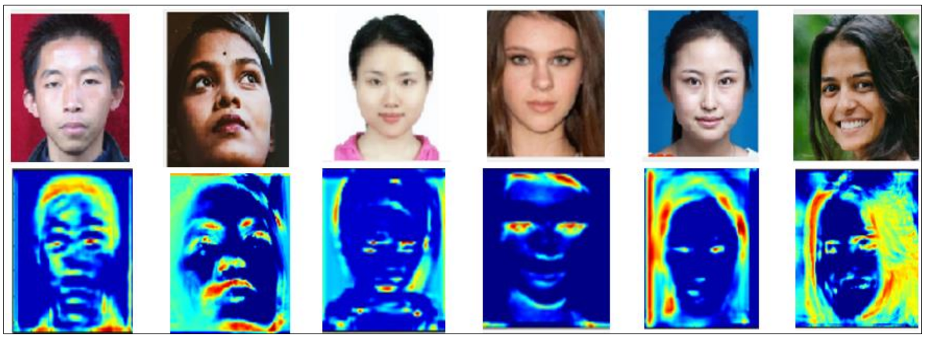


Figure 10. Samples of applying the Grad-CAM analyzing technique

Table 5. The performance comparison with state-of-the-art

	SCUT-FBP			SCUT-FBP 5500			ME Beauty		
Method	↑PC	↓MAE	↓RMSE	↑PC	↓MAE	↓RMSE	↑PC	↓MAE	↓RMSE
[23]	0.8570	0.2595	0.3397	A/N	A/N	A/N	A/N	A/N	A/N
[20]	0.8977	A/N	A/N	0.8783	0.2501	0.3263	A/N	A/N	A/N
[21]	A/N	A/N	A/N	0.885	0.243	0.326	0.748	0.674	0.877
Ours	0.879542	0.226	0.330	0.886	0.242	0.320	0.888	0.365	0.600

5. CONCLUSIONS

CNN has made considerable progress in computer vision in recent years due to its incredible capacity to learn discriminative features. To get the merits of the notion of training a deep network from scratch and the idea of transferring the knowledge of the pre-trained networks, an ensemble DCNNs regression-based model for automatically assessing facial beauty is proposed. The primary goal of the suggested model is to reach a decision relying on the average performance of three different FBP models that have been acquired using different datasets and distinct learning environments. The results are promising, showing the efficiency of fusing the decision of multiple predictors of the proposed ensemble DCNNs regression model, which is noticeably consistent with the ground truth of the employed datasets, and it can help in understanding the relationship between facial characteristics and the impression of attractiveness through evaluating the most discriminating features found by CNN using the Grad-CAM technique. The suggested method has several potential applications, such as pre-and post-operative evaluations in plastic surgery and assessing facial image beautification.

REFERENCES

[1] Saeed, J., Abdulazeez, A.M. (2021). Facial beauty prediction and analysis based on deep convolutional neural network: A review. Journal of Soft Computing and Data Mining, 2(1): 1-12.

[2] Zhang, D., Zhao, Q., Chen, F. (2011). Quantitative analysis of human facial beauty using geometric features. Pattern Recognition, 44(4): 940-950. https://doi.org/10.1016/j.patcog.2010.10.013

[3] Eisenthal, Y., Dror, G., Ruppin, E. (2006). Facial attractiveness: Beauty and the machine. Neural Computation, 18(1): 119-142. https://doi.org/10.1162/089976606774841602

[4] Zebari, G.M., Zebari, D.A., Zeebaree, D.Q., Haron, H., Abdulazeez, A.M., Yurtkan, K. (2021). Efficient CNN Approach for Facial Expression Recognition. In Journal of Physics: Conference Series, 2129(1): 012083. https://doi.org/10.1088/1742-6596/2129/1/012083

[5] Abdullah, S.M.S., Abdulazeez, A.M. (2021). Facial expression recognition based on deep learning convolution neural network: A review. Journal of Soft Computing and Data Mining, 2(1): 53-65.

[6] He, K., Girshick, R., Dollár, P. (2019). Rethinking imagenet pre-training. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 4918-4927.

[7] Liu, S., Fan, Y.Y., Samal, A., Guo, Z. (2016). Advances in computational facial attractiveness methods. Multimedia Tools and Applications, 75: 16633-16663. https://doi.org/10.1007/s11042-016-3830-3

[8] Siqueira, H., Magg, S., Wermter, S. (2020). Efficient facial feature learning with wide ensemble-based convolutional neural networks. In Proceedings of the AAAI Conference on Artificial Intelligence, 34(4): 5800-5809. https://doi.org/10.1609/aaai.v34i04.6037

[9] Chen, L., Shakhnarovich, G. (2014). Learning ensembles of convolutional neural networks. The University of Chicago.

[10] Fan, J., Chau, K. P., Wan, X., Zhai, L., Lau, E. (2012). Prediction of facial attractiveness from facial proportions. Pattern Recognition, 45(6): 2326-2334. https://doi.org/10.1016/j.patcog.2011.11.024

[11] Chen, F., Xu, Y., Zhang, D. (2014). A new hypothesis on facial beauty perception. ACM Transactions on Applied Perception (TAP), 11(2): 1-20. https://doi.org/10.1145/2622655

[12] Chen, F., Xiao, X., Zhang, D. (2016). Data-driven facial beauty analysis: Prediction, retrieval and manipulation. IEEE Transactions on Affective Computing, 9(2): 205-216. https://doi.org/10.1109/TAFFC.2016.2599534

[13] Gan, J.Y., Wang, B., Xu, Y. (2015). A novel method for predicting facial beauty under unconstrained condition. In International Conference on Image and Graphics, pp. 350-360.

[14] Yan, M., Duan, Y., Deng, S., Zhu, W., Wu, X. (2016). Facial beauty assessment under unconstrained conditions. In 2016 8th International Conference on Electronics, Computers and Artificial Intelligence (ECAI), pp. 1-6. https://doi.org/10.1109/ECAI.2016.7861087

[15] J Iyer, T., Nersisson, R., Zhuang, Z., Joseph Raj, A.N., Refayee, I. (2021). Machine learning-based facial beauty prediction and analysis of frontal facial images using facial landmarks and traditional image descriptors. Computational Intelligence and Neuroscience, 2021: 4423407. https://doi.org/10.1155/2021/4423407

[16] LeCun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. nature, 521(7553): 436-444.

[17] Gan, J., Xiang, L., Zhai, Y., Mai, C., He, G., Zeng, J., Bai, Z.F., Labati, R.D., Piuri, V., Scotti, F. (2020). 2M BeautyNet: Facial beauty prediction based on multi-task transfer learning. IEEE Access, 8: 20245-20256. https://doi.org/10.1109/ACCESS.2020.2968837

[18] Lebedeva, I., Guo, Y., Ying, F. (2021). Transfer learning adaptive facial attractiveness assessment. In Journal of Physics: Conference Series, 1922(1): 012004. https://doi.org/10.1088/1742-6596/1922/1/012004

[19] Anderson, R., Gema, A.P., Isa, S.M. (2018). Facial attractiveness classification using deep learning. In 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), pp. 34-38. https://doi.org/10.1109/INAPR.2018.8627004

[20] Xu, L., Fan, H., Xiang, J. (2019). Hierarchical multi-task network for race, gender and facial attractiveness recognition. In 2019 IEEE International Conference on Image Processing (ICIP), pp. 3861-3865. https://doi.org/10.1109/ICIP.2019.8803614

[21] Lebedeva, I., Guo, Y., Ying, F. (2021). MEBeauty: A multi-ethnic facial beauty dataset in-the-wild. Neural Computing and Applications, 1-15. 10.1007/s00521-021-06535-0

[22] Saeed, J.N., Abdulazeez, A.M., Ibrahim, D.A. (2022). FIAC-Net: Facial image attractiveness classification based on light deep convolutional neural network. In 2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA), pp. 1-6. https://doi.org/10.1109/ICCSEA54677.2022.9936582

[23] Xu, L., Xiang, J., Yuan, X. (2018). Transferring rich deep features for facial beauty prediction. arXiv preprint arXiv:1803.07253. https://arxiv.org/abs/1803.07253

[24] Beyer, L., Hénaff, O.J., Kolesnikov, A., Zhai, X., Oord, A.V.D. (2020). Are we done with imagenet? arXiv preprint arXiv:2006.07159.

https://arxiv.org/abs/2006.07159

[25] Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[26] Krizhevsky, A., Sutskever, I., Hinton, G.E. (2017). Imagenet classification with deep convolutional neural networks. Communications of the ACM, 60(6): 84-90. https://doi.org/10.1145/3065386

[27] Motepe, S., Hasan, A.N., Shongwe, T. (2022). Forecasting the total South African unplanned capability loss factor using an ensemble of deep learning techniques. Energies, 15(7): 2546. https://doi.org/10.3390/en15072546

[28] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In Proceedings of the IEEE International Conference on Computer Vision, pp. 618-626.

[29] Liu, C., Meerten, Y., Declercq, K., Gryllias, K. (2022). Vibration-based gear continuous generating grinding fault classification and interpretation with deep convolutional neural network. Journal of Manufacturing Processes, 79: 688-704. https://doi.org/10.1016/j.jmapro.2022.04.068

[30] Fernandes, A.M., Utkin, A.B., Chaves, P. (2022). Automatic early detection of wildfire smoke with visible light cameras using deep learning and visual explanation. IEEE Access, 10: 12814-12828. https://doi.org/10.1109/ACCESS.2022.3145911

[31] Xie, D., Liang, L., Jin, L., Xu, J., Li, M. (2015). Scut-fbp: A benchmark dataset for facial beauty perception. In 2015 IEEE International Conference on Systems, Man, and Cybernetics, pp. 1821-1826. https://doi.org/10.1109/SMC.2015.319

[32] Liang, L., Lin, L., Jin, L., Xie, D., Li, M. (2018). Scut-fbp5500: A diverse benchmark dataset for multi-paradigm facial beauty prediction. In 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, pp. 1598-1603. https://doi.org/10.1109/ICPR.2018.8546038

[33] Maulud, D., Abdulazeez, A.M. (2020). A review on linear regression comprehensive in machine learning. Journal of Applied Science and Technology Trends, 1(4): 140-147. https://doi.org/10.38094/jastt1457