applied sciences

MDPI

Article

Automatic Facial Aesthetic Prediction Based on Deep Learning with Loss Ensembles

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

[1] Technical Informatics College of Akre, Duhok Polytechnic University, Duhok 42001, Iraq
[2] Technical College of Engineering-Duhok, Duhok Polytechnic University, Duhok 42001, Iraq; adnan.mohsin@dpu.edu.krd
[3] IT Collage, Imam Ja'afar AlSadiq University, Baghdad 10001, Iraq; dheyaa.ibrahim@sadiq.edu.iq
[*] Correspondence: jwan.najeeb@dpu.edu.krd

Abstract: Deep data-driven methodologies have significantly enhanced the automatic facial beauty prediction (FBP), particularly convolutional neural networks (CNNs). However, despite its wide utilization in classification-based applications, the adoption of CNN in regression research is still constrained. In addition, biases in beauty scores assigned to facial images, such as preferences for specific, ethnicities, or age groups, present challenges to the effective generalization of models, which may not be appropriately addressed within conventional individual loss functions. Furthermore, regression problems commonly employ L2 loss to measure error rate, and this function is sensitive to outliers, making it difficult to generalize depending on the number of outliers in the training phase. Meanwhile, L1 loss is another regression-loss function that penalizes errors linearly and is less sensitive to outliers. The Log-cosh loss function is a flexible and robust loss function for regression problems. It provides a good compromise between the L1 and L2 loss functions. The Ensemble of multiple loss functions has been proven to improve the performance of deep-learning models in various tasks. In this work, we proposed to ensemble three regression-loss functions, namely L1, L2, and Log-cosh, and subsequently averaging them to create a new composite cost function. This strategy capitalizes on the unique traits of each loss function, constructing a unified framework that harmonizes outlier tolerance, precision, and adaptability. The proposed loss function's effectiveness was demonstrated by incorporating it with three pretrained CNNs (AlexNet, VGG16-Net, and FIAC-Net) and evaluating it based on three FBP benchmarks (SCUT-FBP, SCUT-FBP5500, and MEBeauty). Integrating FIAC-Net with the proposed loss function yields remarkable outcomes across datasets due to its pretrained task of facial-attractiveness classification. The efficacy is evident in managing uncertain noise distributions, resulting in a strong correlation between machine- and human-rated aesthetic scores, along with low error rates.

Keywords: facial beauty prediction (FBP); knowledge transfer; regression-based CNN; Log-cosh loss function; ensemble learning

check for updates

Citation: Saeed, J.N.; Abdulazeez, A.M.; Ibrahim, D.A. Automatic Facial Aesthetic Prediction Based on Deep Learning with Loss Ensembles. *Appl. Sci.* **2023**, *13*, 9728. https://doi.org/10.3390/app13179728

Academic Editor: Krzysztof Koszela

Received: 6 June 2023
Revised: 12 August 2023
Accepted: 14 August 2023
Published: 28 August 2023

1. Introduction

The beauty-related industries have witnessed significant expansion globally, owing to their multifarious beneficial applications in the entertainment, digital media, plastic surgery, and cosmetic sectors. To this end, various studies have been conducted by researchers, medical practitioners, and artists to investigate and measure facial beauty [1–3]. The task of predicting facial attractiveness is a complex and important undertaking in the field of computer vision and machine learning. Building robust and effective FBP models is challenging due to the variability of facial appearance and the complexity of human perception. Conventional methods for predicting face beauty were based on hand-crafted features; these features are manually extracted and then fed to a classifier or a regressor [4].

Facial geometry, color, texture, and other local characteristics are some examples of feature-based representation. Nonetheless, the effectiveness of statistical and traditional machine-learning techniques for extracting and predicting beauty features diminishes with the emergence of sophisticated deep neural networks [5]. The remarkable capability of CNNs to learn discriminative features has led to significant advancements in computer vision. In addition, both deep and machine learning are characterized by applying statistical methods to allow computers to learn from the data supplied.

The machine-learning predictions error is measured using loss and cost functions. Loss functions quantify the error per observation, whereas cost functions quantify the error across all observations. Therefore, loss functions are the core aspects of the training process in the machine-learning system. The ideal values of parameters are derived by mitigating the mean value of the loss, given a labeled training set; hence, selecting the appropriate loss function is the greatest priority. FBP may assume various forms, including classification, regression, or ranking problems [6]. However, CNN is mostly used in classification-based applications, and its implementation in regression studies is still developing. Thus, predicting a collection of dependent, continuous variables is the focus of the latest suggested deep-learning techniques that tackle the facial-image-attractiveness assessment task. Consequently, choosing a loss function that matches the specific predictive modeling issue, such as classification or regression, is critical so that CNN models can learn from the data.

The mean square error (MSE), which is based on L2 loss, is commonly employed for regression problems, aiming to reduce the squared difference between the predicted and the actual values. However, it is sensitive to outliers, making it hard to generalize based on how many outliers might be present via the training phase [7]. While the mean absolute error (MAE) based on L1 loss can cover the MSE downside as it takes the absolute value into account, the errors will be penalized linearly, and it is less sensitive to outliers [8]. However, with the MAE, all errors are equally weighted. In addition, it is not a differentiable function. A viable solution was proposed in [9] to manage the actual and predicted values when data are susceptible to unknown noise distributions by applying the logarithmic hyperbolic cosine (lncosh) as a loss function. This outcome makes sense, given that the Log-cosh function takes advantage of L2 loss for small values and the L1 loss for large ones. Moreover, the Log-cosh loss function is a type of robust loss function that does not require the adjustment of hyperparameters. This contrasts with other robust loss functions, such as Huber loss [10] and Tukey loss [11], which require tuning hyperparameters for optimal performance. Additionally, the Log-cosh loss function has a smooth gradient, which allows for more efficient optimization and convergence during model training. The Log-cosh loss function provides a simple yet effective solution for robust regression tasks, without extensive hyperparameter tuning.

In machine learning, the bias–variance dilemma emerges from the endeavor to minimize bias and variance simultaneously. Complex models often possess low bias and high variance, whereas relatively simple models tend to have high bias and low variance. This predicament encapsulates the challenge of enabling a model to grasp the optimal input-output relationship while maintaining the capacity to generalize beyond the original training data samples. Model ensembles, much like voting systems, involve each member contributing equally through predictive votes [12]. This incorporation of varied perspectives fosters diversity within the ensemble, ultimately leading to diminished variance and enhanced generalization capabilities beyond training data. The voting regressor utilizes the ensembling concept by averaging individual predictions for a conclusive outcome.

Facial-beauty data may exhibit biases, which can affect data distribution. For example, there may be a bias towards images of certain beauty scores, ethnicities, or age groups which can affect the generalizability of the model trained on the dataset. The integration of multiple loss functions in a linear combination has proven to enhance results, as supported by previous studies [13–15]. Thus, the most effective solution in addressing these biases and ensuring the generalizability of the model is using a combination of loss functions to integrate diverse data-driven techniques and leverage their benefits. This work represents a pioneering effort at incorporating the Log-cosh loss function into FBP, combined with L1 and L2 losses, using ensemble average in a CNN-regression-based model for predicting facial-image beauty scores. This distinctive contribution highlights the originality and uniqueness of our work in the field of FBP through ensembling the average of L2, L1, and Log-cosh regression-loss functions within a deep network. Accordinhgly, the model can better capture the underlying patterns and relationships in the data, leading to improved generalization and robustness in FBP across various demographic factors, such as gender, age, and ethnicity. Recognizing the significance of improving prediction reliability and precision, the fusion of diverse data-driven techniques into an ensemble has become a prominent research domain in recent times. The key contributions of this paper are as follows:

1. Leveraging the Log-cosh loss function within the context of FBP to enhance the learning process; to the best of our knowledge, we are the first to use it in quantifying the beauty in facial images.
2. Refining the performance of three distinct pretrained CNNs, namely AlexNet, VGG16, and FIAC-Net, for the purpose of estimating the beauty score within facial images. This enhancement is achieved through the process of tuning and retraining these networks on separate regression-loss functions, namely L1 loss, L2 loss, and Log-cosh, on an individual basis for each network.
3. Proposing a new ensemble average cost function that effectively combines L2, L1, and Log-cosh loss functions to enhance the model's generalization and robustness in predicting the beauty scores across subjects with diverse gender, age, and ethnic characteristics, and further integrating this ensemble cost function with diverse CNN models demonstrate its efficacy in enhancing the capabilities of various deep-learning architectures for the FBP task.
4. Utilizing distinct FBP benchmarks with the aim of comparing the performance of the proposed models in both restricted environments, represented by the SCUT-FBP and SCUT-FBP5500 datasets, and wild facial images captured under unconstrained conditions, as represented by the MEBeauty dataset.

This paper is structured in a manner that includes a brief overview of the relevant studies in Section 2, followed by a description of the suggested framework in Section 3. The empirical findings are presented in Section 4, while Section 5 provides a concluding summary of the study.

2. Related Work

FBP has received significant attention within computer vision as an emerging research area. Estimating the beauty level from a facial image could be treated as a classification [16–18], regression [19,20], or ranking [21,22] task. Two distinct categories of FBP exist. The first category employs a combination of hand-crafted features and conventional machine learning, whereas CNN and deep-learning techniques facilitate the second. Earlier research on FBP focused mostly on a set of hand-crafted features (geometric and texture) that led to the shallow machine-learning algorithms used to estimate facial aesthetics. Hong et al. [23] considered a set of facial ratios as an objective of facial beauty criteria to be incorporated into ensemble-regression-based predictors to obtain the beauty score. However, geometric-feature-based techniques have limited performance due to the influence of facial-expression variation, and it demands a computational burden through landmark localizations. Psychological studies confirm that facial color, smoothness, and lightness are crucial for perceiving facial beauty [24–26]. Iyer et al. [27] implemented conventional image descriptors for texture feature extraction and combined them with putative facial ratios to predict the attractiveness score. In their experiments, the K-nearest neighbors (KNNs) achieved the highest performance for the combined features, with a Pearson correlation of 0.7836 and MSE of 0.0963, to outperform the other suggested models, such as linear regression, random forest, and artificial neural network (ANN), when evaluated on SCUT-FBP5500 dataset.

Feature engineering was once considered a crucial part of computer vision applications before the emergence of deep learning. Recently, there has been an increasing demand for the automatic extraction of features from facial images. CNNs provide an end-to-end learning approach that can learn the mapping from the input to the desired output, eliminating the need for manual feature engineering. An ensemble CNN-based regression model was proposed in [28] to estimate the facial-beauty score automatically by utilizing a diverse learning environment and fuse the decision made by these ensembled networks to obtain a more reliable aesthetic score. However, sometimes the dominance of people with average beauty scores in the facial-beauty data space may pose a challenge when optimizing models using traditional loss functions, potentially resulting in extremely attractive and unattractive faces being treated as outliers.

It is imperative to employ a hybrid loss function tailored based on the estimated tasks to predict multiple distinct but beauty-related tasks efficiently. Consequently, researchers have begun to explore models capable of handling outliers and multiple tasks simultaneously, resulting in improved performance compared to conventional regression approaches that rely heavily on L1 and L2 loss functions [29,30]. For instance, Gao et al. [31] utilized multiple loss functions for a CNN that simultaneously performed facial landmark localization and FBP tasks. Furthermore, a multitask CNN named HMTNet was introduced in [32]. HMTNet can predict the attractiveness score of a facial image, along with its race and gender. They devised a loss function called "Smooth Huber Loss". Similarly, Lebedeva et al. [33] employed the Huber loss function to assess various known pretrained CNNs on their recently introduced multi-ethnic dataset, as they anticipated a substantial number of outlier faces. However, the choice of the delta parameter in Huber loss can significantly impact the optimization process and ultimately the prediction accuracy, which may require additional hyperparameter tuning. Zhai et al. [34] developed an effective CNN based on the pretrained face-recognition model (CASIA-WebFace) and utilized a double activation layer with a new Softmax-MSE loss function to predict the facial beauty in Asian females. In the same context, Lu et al. [15] introduced a new cost function by incorporating a weighted cross-entropy loss for classification, along with an expectation loss and a regression loss called ComboLoss to direct the SEResNeXt50 FBP network to improve model training. However, their combination may increase computational complexity and difficulty in tuning the hyperparameters. More recently, Dornaika et al. [35] suggested a two-branch net (REX-INCEP) based on the ResneXt-50 and Inceptionv3 structures, and then the ParamSmoothL1 regression-loss function was developed to estimate face beauty. However, the dynamic parameterized smoothL1 loss function can be computationally expensive compared to other loss functions. This is because it involves a piecewise function that requires additional calculations, which can slow down the training process.

3. Methodology

The general framework of the proposed method is depicted in Figure 1. First, the input data were partitioned into training and testing folds based on 5-fold cross-validation. The training data were then fed to a deep network to learn the pattern of estimating the beauty score in a facial image. Then, the network was trained on the proposed cost function that ensembles three regression-loss functions: L2 loss, L1 loss, and Log-cosh loss. This ensemble objective function can improve the learning efficiency by inheriting the benefits of each loss and minimizing the loss value to validate the model performance. Ultimately, the trainable model was evaluated on the test data to show the model efficiency. The next subsections elaborate on these stages in further detail.

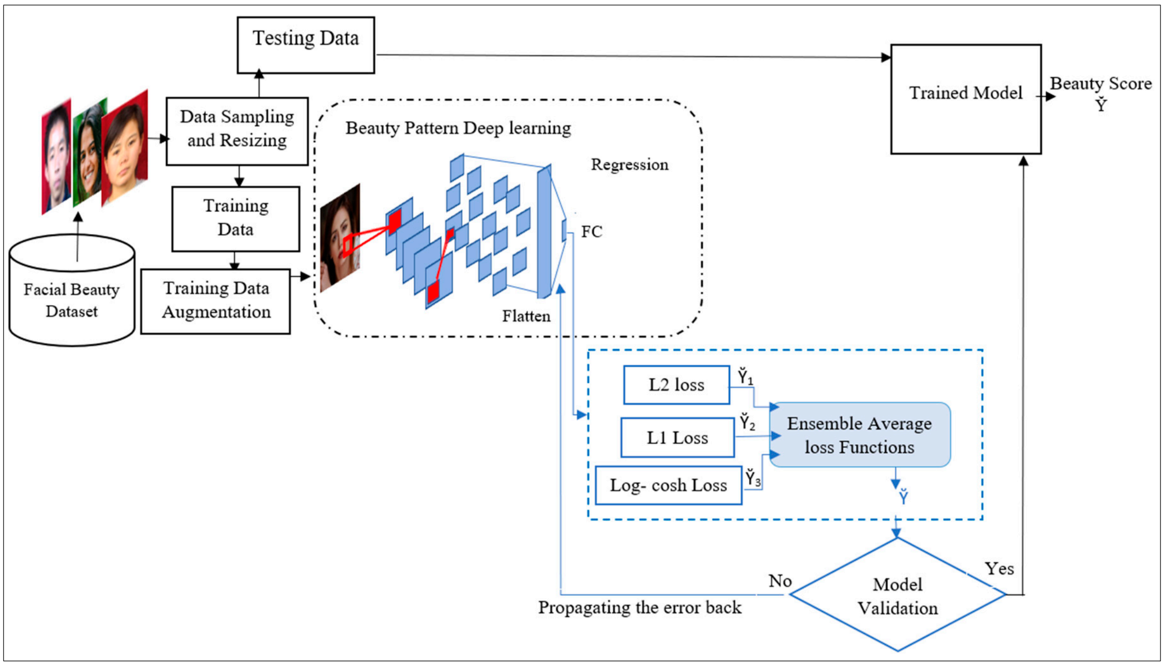

Figure 1. Schematic representation of the proposed model.

3.1. Image Preprocessing

FBP datasets have a variety of image sizes that are required to be resized according to the size of the network's input layer. Moreover, the limited sample size and unbalanced data present a concern that may lead to an overfitting problem. Consequently, partitioning data samples based on a cross-validation technique can help to avoid overfitting by providing a more accurate estimate of how well the model works on data it has not seen before. This work employed the 5-fold cross-validation procedure, which usually involves randomly separating all the data into five folds. The model is then trained on four folds (80%), and the remaining fold is utilized to test the model (20%).

The SCUT-FBP was subjected to a stratified partitioning scheme with five folds. Each fold consisted of 400 images allocated for training purposes, while 100 images were reserved for testing. Similarly, the SCUT-FBP5500 dataset underwent the same partitioning process, resulting in five folds with 4400 images for training and 1100 images for testing. Furthermore, the MEBeauty dataset was subjected to a similar partitioning strategy, where each of the five folds contained 2040 training images and 510 testing images. This rigorous partitioning scheme ensured a systematic evaluation of the models' performance on distinct subsets of the datasets.

3.2. The Loss Functions

A crucial component of a learning system is a loss function that quantifies the accuracy of the predicted value relative to the ground-truth value. In machine learning, there are two loss functions: those based on the margin employed in the classification process and those based on distance in regression problems [36]. The loss function is denoted as follows:

$$\text{loss function} = L(\vec{x},\ y_t;\ y_p) \tag{1}$$

$$y_p \ = \ f(\vec{w}^{T}.\ \vec{x}) \tag{2}$$

where $\vec{x}$ denotes the input, y_t is the target (actual) value, and y_p is the predicted value. The weight in the network is represented by w. Meanwhile, $f()$ refers to the learning algorithm.

The smaller loss value indicates the model's efficiency. The choice of training loss function considerably impacts the model's performance and generalizability [37]. More detail regarding the regression-loss functions used in this study is presented in the subsequent subsection. The pertinent variables and notations employed in the proposed methodology are summarized in Table 1.

Table 1. The description of the pertinent variables and notations.

Parameter	Description
$\vec{x}$	Input image
y_t	Target ground truth of beauty score
$\overline{y}$	The average of the ground-truth scores
y_p	Predicted beauty score
$\overline{p}$	The average of the predicted scores
n	Number of data samples
$f()$	Learning function or algorithm
w	Weight of the input data
L	Loss function
j	Number of loss functions
L1	L1 loss
L2	L1 loss
Log-cosh	Hyperbolic cosine h

3.2.1. L2 Loss

The regression theorem and least-squares approach are the origins of this function. It is the sum of the squared distances between the target value y_t and the predicted value y_p. The L2 loss formulation is presented as:

$$L2 = \left(y_t - y_p\right)^2 \tag{3}$$

3.2.2. L1 Loss

L1 loss refers to the average of the absolute differences between actual and predicted values. It is more effective as a loss function when the data have several outliers. Due to the squaring of errors, even a few outliers can have a huge impact on the L2 loss, leading to overestimating the errors associated with the outliers. While L1 loss is a linear function that gives equal weight to all deviations, it is less responsive to outliers than L2 since it does not penalize the high deviations caused by the outliers. The formula of L1 is shown as follows:

$$L1 = \left|y_t - y_p\right| \tag{4}$$

3.2.3. Log-cosh Loss

The Log-cosh is the hyperbolic cosine of the difference between the predicted value, y_p, and the true value, y_t. The notion of the Log-cosh loss is described as follows:

$$Log\_cosh \ = \ log(cosh(y_{\ p} - y_t)) \tag{5}$$

where we have

$$\cosh(y_p - y_t) = \frac{e^{(y_p - y_t)} + e^{-(y_p - y_t)}}{2} \tag{6}$$

More specifically,

$$Log\_cosh\left(y_p - y_t\ \right) = log(\frac{e^{(y_p - y_t)} + e^{-(y_p - y_t)}}{2})$$

The Log-cosh function is smoother than quadratic loss. It works like L2, but it is not affected by substantial prediction errors, and it also mimics a smoothed version of the L1 loss that is differentiable everywhere [38].

3.3. The Proposed Ensemble Cost Function

Ensemble learning is a widely recognized approach that involves combining the predictions of multiple learning models to improve the overall performance [39]. Using three diverse loss functions balances individual strengths and weaknesses. Averaging these losses can enhance the model's ability to address a wider range of prediction errors. For instance, if one loss function is sensitive to outliers, including others with distinct characteristics can mitigate their impact on optimization. This strategy promotes a more holistic optimization landscape, potentially improving the generalization and robustness of predictions. To identify patterns for prediction tasks, discriminative deep-learning models are employed for supervised learning. This article introduces a new ensemble cost function that leverages three distinct loss functions, namely, L1 loss, L2 loss, and Log-cosh loss, to enhance the performance of CNN in regression tasks. The proposed ensemble-learning approach exhibits a key strength in enabling the effective fusion of multiple sources of information. Moreover, the method offers a robust solution for handling unknown noise distributions commonly encountered in real-world applications.

Based on Equations (1) and (2), the aggregation of the proposed ensemble average loss functions is formulated as follows:

$$\mathrm{Ensemble\ \ cos\,t\ function} = \frac{1}{3n}\sum\nolimits_{i=1}^{n}\sum\nolimits_{j=1}^{3} L_j(\vec{x}, y_t^i; y_p^i) \tag{7}$$

where n is the number of data samples, L_j represents the loss function, and j denotes the number of loss functions. It could be formulated more specifically as follows:

$$\mathrm{Ensemble\ \ cos\,t\ function} = \frac{1}{3n}\sum\nolimits_{i=1}^{n}\left|y_t^i - y_p^i\right| + \left(y_t^i - y_p^i\right)^2 + (log(cosh(y_p^i - y_t^i)))$$

The computational cost and time associated with utilizing multiple loss methods warrant consideration. While a single loss function might appear to be computationally efficient, combining multiple functions could introduce added complexity due to calculating and merging multiple loss terms. Nevertheless, potential performance gains might offset the heightened computational demands.

3.4. Beauty Pattern Deep Learning and Knowledge Transfer

Transfer learning is a recent deep-learning baseline that overcomes overfitting when training samples are limited. Furthermore, the pretrained networks are regarded as important in machine-learning-model evaluation. To show the effectiveness of the proposed loss function in estimating the beauty score of the facial image, three pretrained CNNs (AlexNet [35], VGG16-Net [36], and FIAC-Net [37]) were fine-tuned to obtain benefit from the gained knowledge of each network, utilizing the transfer-learning aspects.

In theory, deeper networks should outperform shallower ones. However, in practice, deeper networks tend to have higher computational complexity and can be susceptible to overfitting, especially when trained on relatively small datasets. Consequently, the CNN models utilized in this work adopt a moderate layer configuration. Notably, both AlexNet and VGG16-Net have demonstrated efficacy across diverse computer vision tasks due to their pretraining on the large-scale ImageNet dataset [40] to recognize and classify objects across thousands of different categories.

Meanwhile, FIAC-Net [41] is a lightweight CNN that is pretrained to classify attractiveness levels in facial images, using FBP datasets, including the CelebA dataset [42], which encompasses more than 200K facial images of celebrities. The selection of these CNNs represents a comprehensive approach considering performance, architecture complexity, and their original training task in computer vision.

Fine-tuning the pretrained network involves adjusting the hyperparameters of a CNN, including the number of layers, neurons, epochs, learning rate, and the associated cost function. Additionally, it improves the performance of the training model to achieve the finest performance accuracy and aims to obtain the optimal set of parameters. The proposed methodology fine-tunes these networks to be more suitable for the facial-photo beauty-estimation problem. Moreover, to address the overfitting concern, training-data augmentation is implemented. This involves augmenting the training images with random rotations, translations, and reflections.

To train the utilized CNNs, the Adam optimizer was utilized, and the number of training epochs was set to 150, with a batch size of 32. Furthermore, the initial learning rate was established at 1×10^{-4}. Table 2 depicts the training hyperparameters' setting details.

Table 2. The hyperparameters' configuration for model training.

Parameter	Setting
Initial learning rate	1×10^{-4}
Batch size	32
MaxEpochs	150
Optimizer	Adam

The pseudocode represented in Algorithm 1 outlines the implementation of knowledge transfer in the proposed models which leverages a pretrained network to enhance the performance of an FBP task. By freezing certain layers, modifying the network architecture, and applying regression-loss functions, the model learns to predict beauty scores from facial images.

3.4.1. FBP Based on AlexNet

The AlexNet [43] is an eight-layer CNN comprising five convolutional layers and three fully connected layers. Figure 2 and Table 3 show the layer configurations of the proposed AlexNet-regression-based for FBP. The ReLU activation function is employed in all layers except for the output layer that uses the Softmax activation function. To accurately estimate the facial attractiveness score, the knowledge of AlexNet was transferred, and its architecture was modified. This involved substituting the regression-loss function for Softmax, which transformed the classification process into a regression task. Moreover, this was due to the fact that the lower layers of a CNN can learn general low-level features, while higher layers capture task-specific features. We proposed reusing the pretrained weights of lower layers and tuning higher ones. This approach saves time, lets the network adapt efficiently, prevents overfitting, and preserves generalization ability; it was applied to freeze the first two convolutional layers and tune the rest of the layers to be adapted for facial aesthetic assessment.

Table 3. The configurations of the proposed AlexNet-regression-based for FBP.

Layer Name	Kernels	Size	Stride
Input		227 × 227 × 3	
Convolutional_1 + BN + ReLU	96	11 × 11	4
Max pooling_1		3 × 3	2
Convolutional_2 + BN + ReLU	256	5 × 5	1
Max pooling_2		3 × 3	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		3 × 3	2
Fully Connected fc6			
Fully Connected fc7			
Fully Connected fc8			
The proposed ensemble loss with response			
Regression			

Algorithm 1: The pseudocode of the knowledge-transfer implementation for the proposed models

```
Input: Training, validation sample sets: ($X$ $train$, $y$ $train$), and ($X$ $val$idation, $y$ $val$idation).
Output: Predicted facial beauty score r: $y$.

•   Start.
1. Load the pretrained network:
    net = pretrainedNetwork;
2. Freeze the required number of layers based on the network's architecture and its initial task design:
numFrozenLayers = desiredNumFrozenLayers;
        for i = 1:numFrozenLayers
            net.Layers(i).Trainable = false;
        End.
3. Remove fully connected layers:
            net = removeLayers(net, {'ClassificationLayer_softmax', 'fc'}).
4. Replace Softmax and classification layer with a regression-loss function:
        outputLayer = regressionLayer('Name', 'regressionLayer').
5. Add a new fully connected layer for beauty score prediction:
        numOutputNodes = 1;
        fc = fullyConnectedLayer(numOutputNodes, 'Name', 'fc').
6. Adjust hyperparameters and training options.
7. Train the model with 5-fold cross-validation (tuning the rest of the unfrozen layers):
        numFolds = 5;
        partitionedData = cvpartition(numSamples, 'KFold', numFolds);
        for fold = 1:numFolds
            trainingIdx = training(partitionedData, fold);
            trainingDataFold = augmentedData(trainingIdx).
                ○   Train the model using different regression-loss functions individually, namely
                    L1 loss, L2 loss, Log-cosh, and the proposed ensemble average loss
                    function:trainedModel = trainNetwork(net, trainingDataFold, layers(fc,
                    outputLayer), options);
                ○   Save the trained model and evaluation metrics for this fold.
            End.
8. Evaluate model performance on testing data:
            predictions = predict(trainedModel, validationData);
            pc = calculatePearsonCorrelation(predictions, validationScores);
            mae = calculateMeanAbsoluteError(predictions, validationScores);
            mse = calculateMeanSquaredError(predictions, validationScores).
•   Stop.
```

In this work, AlexNet was trained separately, using each of the three loss functions, and the incorporation of the suggested ensemble loss was subsequently examined. The alteration of the AlexNet architecture leads to a more efficient deep-learning network for predicting facial-image attractiveness in regression tasks.

3.4.2. FBP Based on VGG16-Net

The VGG16 network [44] is structured with 3 × 3 convolution layers that extract features, five 2 × 2 max-pooling layers, and three dense layers, as shown in Figure 3 and Table 4. To utilize the VGG16 network in FBP, we fine-tuned it and replaced the final dense layer comprising a large number of neurons with a single neuron to investigate the training of the VGG16 network on each of the three loss functions individually and proposed their combination for the FBP. By incorporating the proposed ensemble loss function into the output layer instead of the Softmax function and freezing the initial three layers, the accuracy and robustness of the fine-tuned model in estimating the facial attractiveness score were enhanced. The alteration of the VGG16 architecture to include the ensemble loss function led to a more efficient performance in the regression task of FBP.

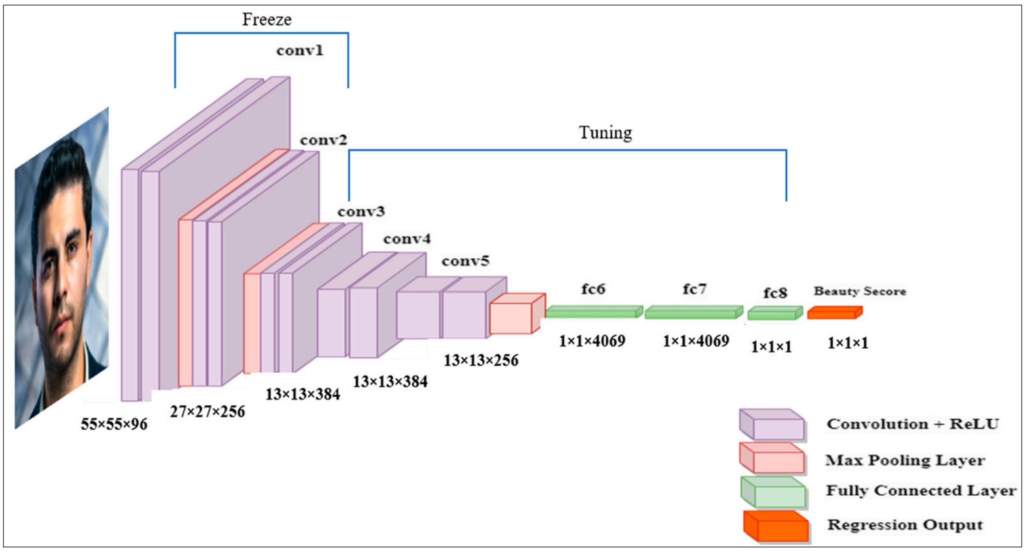

Figure 2. The architecture of the proposed AlexNet-regression-based for FBP.

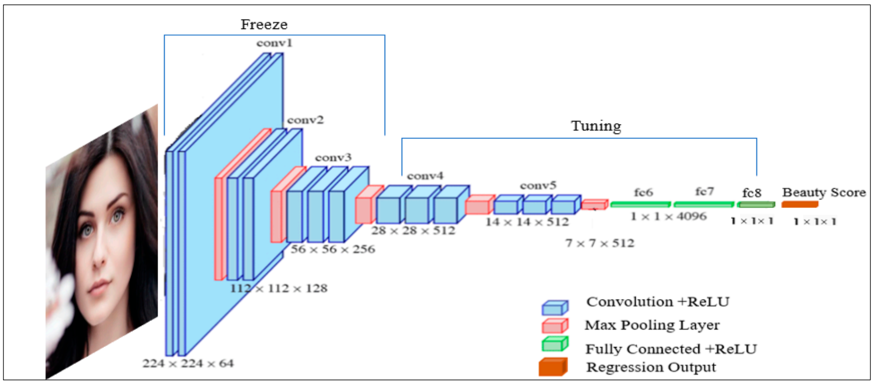

Figure 3. The proposed VGG16-regression-based for FBP.

Table 4. The configurations of the proposed VGG16-Net-regression-based for FBP.

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	3 × 3	1
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

Table 4. *Cont*.

Layer Name	Kernels	Size	Stride
Fully Connected fc6 + ReLU + Dropout			
Fully Connected fc7 + ReLU + Dropout			
Fully Connected fc8			
The proposed ensemble loss with response Regression			

3.4.3. FBP Based on FIAC-Net

FIAC-Net is a lightweight CNN that we developed in [41]. It stands for Facial Image Attractiveness Classification Network. It is constructed with six convolutional layers of varying kernel sizes, accompanied by five layers of 2 × 2 Max pooling. An average-pooling layer precedes the final layer, which is one fully connected layer. The ReLU activation function is uniformly utilized across all layers, with the exception of the output layer that employs the Softmax activation function. FIAC-Net was pretrained to classify the attractiveness level in a facial image by utilizing different FBP benchmarks, including the CelebA dataset. The CelebA dataset [42] contains over 200K facial images of celebrities that represent a diverse range of facial characteristics, poses, and expressions. This variability presents significant challenges for facial-attractiveness classification. This variability allows the deep-learning model to learn and capture informative and robust features related to facial attractiveness. Pretraining on the CelebA dataset can provide a rich source of facial-image data that can aid in learning patterns of facial attractiveness. The FIAC-Net is fine-tuned through the substitution of Softmax with the regression layer that results in the alteration of the classification process into a regression task. Furthermore, the freezing of the initial four convolutional layers is executed, succeeded by the calibration and tuning of the subsequent layers to conform with the prerequisites of FBP adaptation. Figure 4 and Table 5 illustrate the structure of the proposed FBP based on the FIAC-Net model.

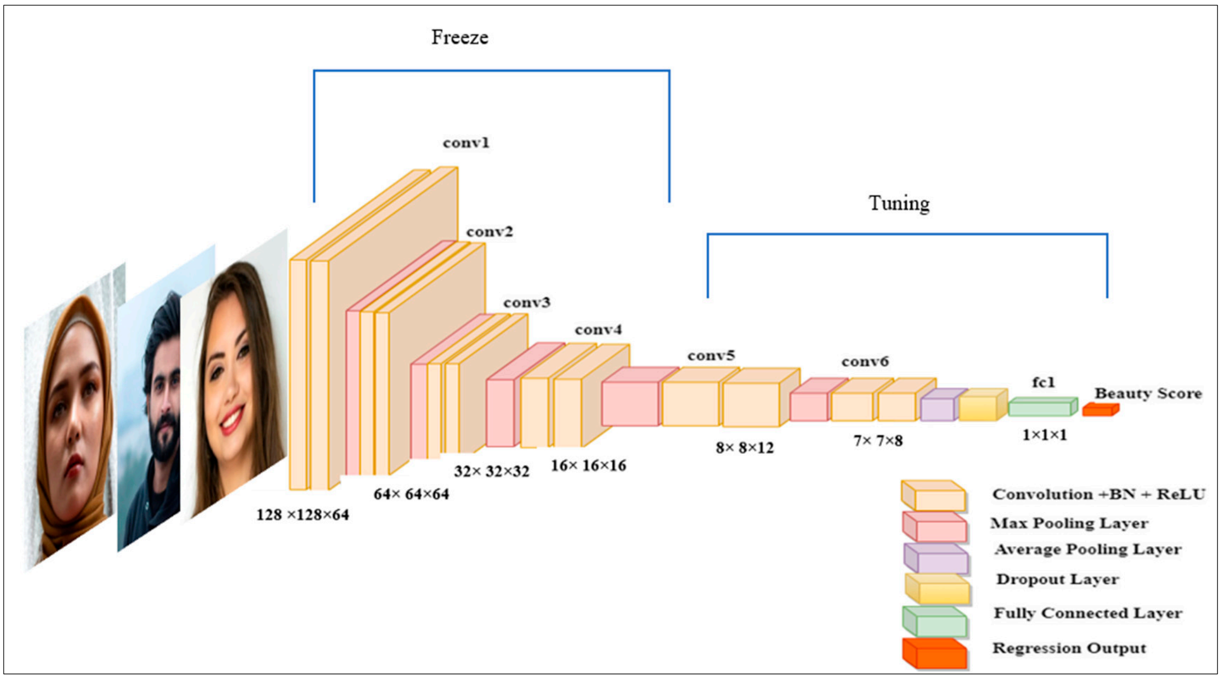

Figure 4. The architecture of the proposed fine-tuned FIAC-Net for FBP.

Table 5. The configurations of the proposed FIAC-Net-regression-based for FBP.

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			
The proposed ensemble loss with response Regression			

4. Experimental Results

This section begins by introducing the evaluation metrics, followed by describing the utilized FBP datasets. Subsequently, individual fine-tuning and retraining of the employed CNNs with different regression-loss functions (L1 loss, L2 loss, Log-cosh loss, and the proposed ensemble loss functions) are highlighted to assess the model's FBP performance. Finally, a comparison is made between the proposed model's findings and the state-of-the-art approaches.

4.1. Metrics of Evaluation

The evaluation of FBP models is mainly based on three metrics: Pearson correlation (PC), mean absolute error (MAE), and root-mean-square error (RMSE). PC ranges from 1 to -1, where 1 indicates a perfect positive linear correlation, 0 denotes no correlation, and -1 represents a perfect negative correlation. MAE and RMSE assess the model's efficiency, with values close to zero indicating a good performance. Given a set of n tests in samples, we have the following:

$$PC = \frac{\sum_{i=1}^{n}\left(y_t^i - \overline{y}\right)\left(y_p^i - \overline{p}\right)}{\sqrt{\sum_{i=1}^{n}\left(y_t^i - \overline{y}\right)^2}\sqrt{\sum_{i=1}^{n}\left(\left(y_p^i - \overline{p}\right)^2\right.}} \tag{8}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_p^i - y_t^i\right| \tag{9}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_p^i - y_t^i\right)^2} \tag{10}$$

where the ground-truth beauty score is represented by y_t^i, and the estimated score for the Ith image is written as y_p^i. Meanwhile, $\overline{y}$ is the average of all of the beauty scores in the ground truth, and $\overline{p}$ is the average of the scores that were predicted by the model. Additionally, a better performance is shown by higher PC values and lower MAE and RMSE values.

4.2. Dataset

The evaluation of integrating the proposed loss into different CNN models was conducted on three prominent FBP benchmarks, encompassing both restricted (SCUT-FBP and SCUT-FBP5500) and unconstrained (MEBeauty) environments. More details on the datasets and evaluation process can be found in the subsequent sections.

4.2.1. SCUT-FBP

SCUT-FBP was created by Xie et al. [45] in 2015. It includes 500 high-quality frontal facial images of Asian females with neutral expressions, simple backgrounds, and little occlusion. The evaluations of attractiveness (scores), which range from 1 to 5, are the outcome of averaging different ratings of 70 raters for each image.

4.2.2. SCUT-FBP5500

The SCUT-FBP5500 [46] dataset comprises a total of 5500 frontal facial images of males and females from both Asian and Caucasian ethnicities, covering a wide range of ages. Each image is assigned a beauty score within the range of 1–5, based on assessments by 60 independent raters. The dataset is evenly split between genders, with 2750 male and 2750 female images. The images were taken in a controlled environment, with uniform lighting and neutral backgrounds, to ensure consistent and accurate capture of facial features.

4.2.3. MEBeauty

MEBeauty is a recently produced multi-ethnic collection comprising 2550 in-the-wild facial photos presented by Lebedeva et al. [33] in 2021. It has the facial images of 1250 men and 1300 women, comprising six ethnic groups, namely Black, Asian, Caucasian, Hispanic, Indian, and Middle Eastern people, with various ages, poses, backgrounds, and expressions represented. The beauty scores within this dataset range 1–10 obtained as an average score of 300 raters per image. Furthermore, the MEBeauty dataset is extremely diverse, with challenging data samples, thus making the beauty quantifying process harder.

4.3. Performance Evaluation and Discussion

The investigation in this work focused on the effectiveness of individually integrating three distinct regression-loss functions—L2, L1, and Log-cosh—along with their ensemble average combination for predicting facial aesthetic scores. The proposed methodology's validation was achieved via a five-fold cross-validation, employing the SCUT-FBP, UT-FBP5500, and MEBeauty datasets, as elaborated upon in subsequent sections.

4.3.1. Performance Evaluation on SCUT-FBP

The distribution of beauty scores in the SCUT-FBP dataset was found to be approximately Gaussian [45,47]. However, this non-perfect normal distribution of beauty scores suggests that relying solely on the traditional L2 loss function may not be optimal in certain scenarios. Table 6 exhibits the performance of three pretrained CNNs on the SCUT-FBP dataset under five-fold cross-validation, using various loss functions. Based on the evaluation metrics, L1 loss produced slightly higher PC values and lower error rates than L2 loss for the three investigated pretrained CNNs. Similarly, the Log-cosh loss function outperformed both the L1 and L2 loss functions among the utilized networks because it provides a more balanced error calculation between large and small residuals. The proposed ensemble loss functions yielded better results for all three fine-tuned CNN models. Specifically, AlexNet achieved a PC value of 0.903758, MAE of 0.26348, and RMSE of 0.348266. On the other hand, VGG16-Net, with more layers than AlexNet, achieved a slightly higher PC value of 0.905851 and lower error rates of 0.222954 and 0.292028 for MAE and RMSE, respectively.

FIAC-Net, unlike pretrained object classification models such as AlexNet and VGG1-Net, was trained specifically for facial aesthetic classification, making it better suited to estimating the beauty scores of faces. The proposed ensemble loss function yielded the best performance with FIAC-Net, achieving a PC value of 0.9100582, the lowest MAE of 0.185949, and an RMSE of 0.259156. Model prediction on test data samples from SCUT-FBP using fine-tuned FIAC-Net is demonstrated in Figure 5. In order to evaluate the effectiveness of the proposed model, Figure 6 visualizes the predictions in a scatter plot. It plots the predicted values on the x-axis against the ground-truth values on the y-axis based on the utilized loss functions, with different pretrained CNNs, namely (a) AlexNet, (b) VGG16-Net, and (c) FIAC-Net, implemented on the SCUT-FBP dataset. The results clearly indicate that the integration of the ensemble loss into FIAC-Net resulted in the most effective scatter-plot representation.

Table 6. Five-fold cross-validation of FBP, assuming different loss functions, using three diverse CNNs on SCUT-FBP.

		AlexNet			**VGG16-Net**			**FIAC-NET**		
Loss	**Fold #**	**PC ↑**	**MAE ↓**	**RMSE ↓**	**PC ↑**	**MAE ↓**	**RMSE ↓**	**PC ↑**	**MAE ↓**	**RMSE ↓**
L2	Fold 1	0.886792	0.309923	0.387405	0.885873	0.233409	0.288079	0.902156	0.159987	0.229707
	Fold 2	0.890809	0.387383	0.461403	0.900324	0.272331	0.364368	0.894374	0.30578	0.462875
	Fold 3	0.894072	0.327179	0.398048	0.896106	0.27479	0.33756	0.900172	0.158858	0.224796
	Fold 4	0.88454	0.277123	0.360177	0.893316	0.286573	0.358064	0.893911	0.272233	0.359504
	Fold 5	0.89616	0.237447	0.35655	0.887568	0.251615	0.349024	0.893666	0.174137	0.232828
	Average	**0.890475**	**0.325402**	**0.392717**	**0.892637**	**0.263744**	**0.339419**	**0.8968558**	**0.214199**	**0.301942**
L1	Fold 1	0.886594	0.268403	0.344794	0.891489	0.188005	0.253954	0.913914	0.179596	0.244818
	Fold 2	0.898972	0.29734	0.399723	0.909594	0.273945	0.334632	0.89788	0.285959	0.426164
	Fold 3	0.897251	0.3151	0.389115	0.89499	0.206798	0.265649	0.901024	0.150231	0.221037
	Fold 4	0.89355	0.291143	0.369973	0.900606	0.28731	0.361998	0.882645	0.250829	0.360831
	Fold 5	0.905704	0.299369	0.398775	0.885381	0.28123	0.396568	0.890992	0.18624	0.244834
	Average	**0.896414**	**0.294271**	**0.380476**	**0.896412**	**0.247458**	**0.32256**	**0.897291**	**0.210571**	**0.299537**
Log-cosh	Fold 1	0.883631	0.313012	0.388109	0.904711	0.236537	0.287591	0.917584	0.168858	0.231028
	Fold 2	0.908497	0.239261	0.343423	0.9077	0.282616	0.352005	0.900933	0.263164	0.345673
	Fold 3	0.89916	0.294613	0.358759	0.901113	0.162906	0.221393	0.90517	0.141968	0.216695
	Fold 4	0.896872	0.256079	0.339267	0.900411	0.2745	0.362309	0.896621	0.246203	0.355604
	Fold 5	0.89441	0.254158	0.369783	0.904146	0.226115	0.303307	0.893336	0.160434	0.236355
	Average	**0.896514**	**0.271425**	**0.359868**	**0.903616**	**0.236535**	**0.305321**	**0.902729**	**0.196125**	**0.277071**
Proposed loss	Fold 1	0.91087	0.277514	0.349938	0.906243	0.177149	0.242459	0.910428	0.16176	0.221287
	Fold 2	0.909427	0.236772	0.339803	0.911665	0.269073	0.344907	0.918538	0.230674	0.315015
	Fold 3	0.899448	0.293994	0.344505	0.903421	0.173957	0.226531	0.913071	0.126127	0.199258
	Fold 4	0.897148	0.297535	0.390668	0.902068	0.258518	0.334192	0.900512	0.259396	0.345144
	Fold 5	0.901899	0.211583	0.316417	0.905859	0.236074	0.312052	0.907742	0.15179	0.215078
	Average	**0.903758**	**0.26348**	**0.348266**	**0.905851**	**0.222954**	**0.292028**	**0.9100582**	**0.185949**	**0.259156**

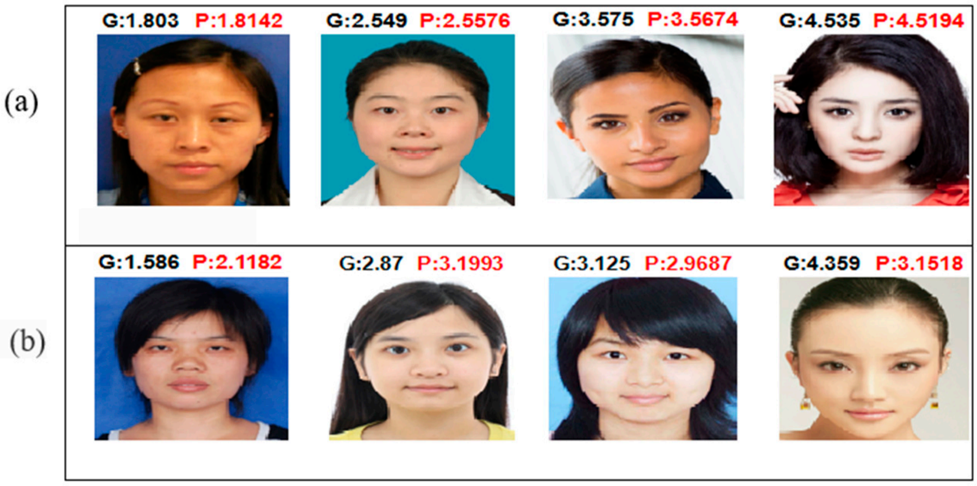

Figure 5. Samples of tested SCUT-FBP data on the fine-tuned FIAC-Net. G, ground truth; P, model-predicted score. (**a**) Accurately predicted instances and (**b**) inaccurately predicted instances.

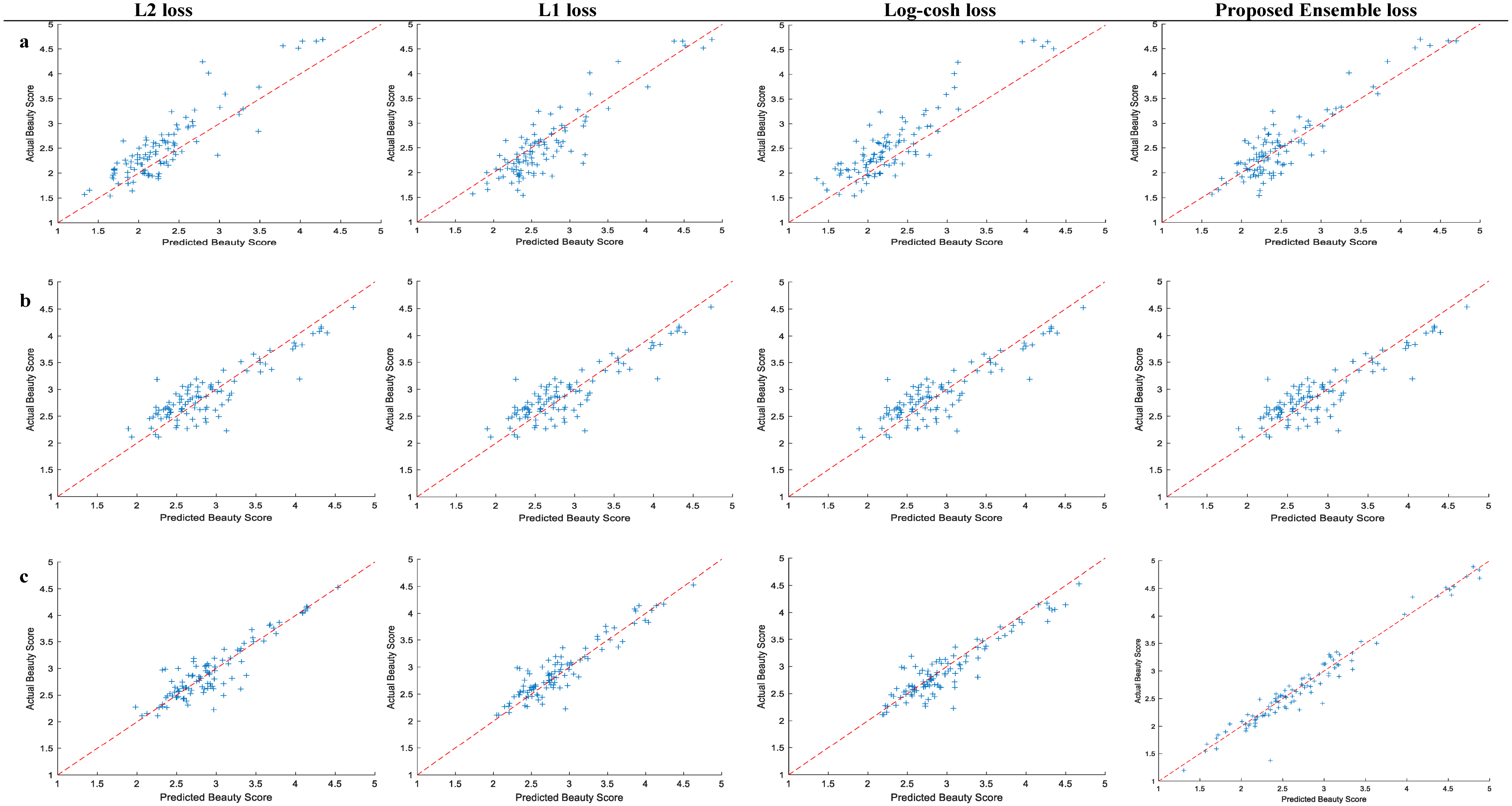

Figure 6. A scatter-plot representation of samples of predicted beauty scores against actual beauty scores on SCUT-FBP: (**a**) AlexNet, (**b**) VGG16-Net, and (**c**) FIAC-Net.

4.3.2. Performance Evaluation on SCUT-FBP5500

Table 7 presents the performance evaluation of three CNNs on the SCUT-FBP5500 dataset. The evaluation was conducted using five-fold cross-validation, and the models were trained with different regression-loss functions. The results indicate that the L2 loss function outperforms the L1 loss, and both L2 loss and Log-cosh loss yield comparable results. This can be attributed to the fact that the data distribution in the SCUT-FBP5500 dataset aligns closely with normal distributions.

When using the L2 loss function, the average performance of AlexNet was a PC of 0.9127168, an MAE of 0.2555128, and an RMSE of 0.3195926. VGG16-Net achieved an average PC of 0.9275814, an MAE of 0.2269836, and an RMSE of 0.2855518. FIAC-Net obtained an average PC of 0.9251664, an MAE of 0.2080304, and an RMSE of 0.2668852.

Meanwhile, the average results for the L1 loss function showed that AlexNet achieved a PC of 0.8944518, an MAE of 0.2473232, and an RMSE of 0.321819. VGG16-Net obtained an average PC of 0.9106684, an MAE of 0.2479496, and an RMSE of 0.3149184. FIAC-Net achieved an average PC of 0.9123388, an MAE of 0.215242, and an RMSE of 0.284987.

When using the Log-cosh loss function, the average performance of AlexNet was a PC of 0.9118102, an MAE of 0.2467066, and an RMSE of 0.3108374. VGG16-Net achieved an average PC of 0.9218162, an MAE of 0.2252112, and an RMSE of 0.2884092. FIAC-Net obtained an average PC of 0.9239984, an MAE of 0.2113572, and an RMSE of 0.2707102.

Furthermore, the proposed ensemble average model, which combined the different loss functions, achieved an average PC of 0.9305098, an MAE of 0.2028174, and an RMSE of 0.2614154. These results indicate that the ensemble model slightly outperformed the L2 loss function in terms of PC, MAE, and RMSE. Consequently, this suggests that the proposed approach can potentially enhance the performance of FBP. Figure 7 illustrates the model predictions obtained by applying the fine-tuned FIAC-Net with the validation fold on the SCUT-FBP5500 dataset. Meanwhile, Figure 8 displays a visual representation of the utilized loss functions incorporated with the fine-tuned CNNs. The results indicate that the proposed ensemble loss function with the fine-tuned FIAC-Net achieves a competitive performance in predicting beauty scores, thus highlighting its efficiency.

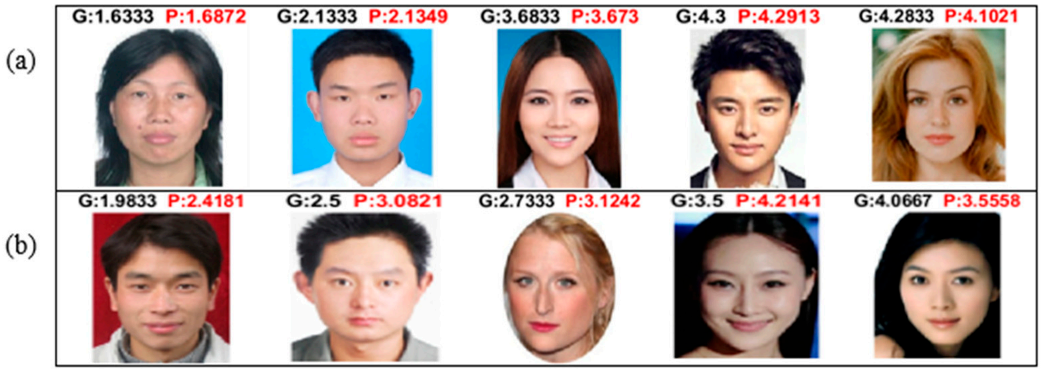

Figure 7. SCUT-FBP5500 tested samples on FIAC-Net. G, ground truth; P, model-predicted score. (**a**) Accurately predicted instances and (**b**) inaccurately predicted instances.

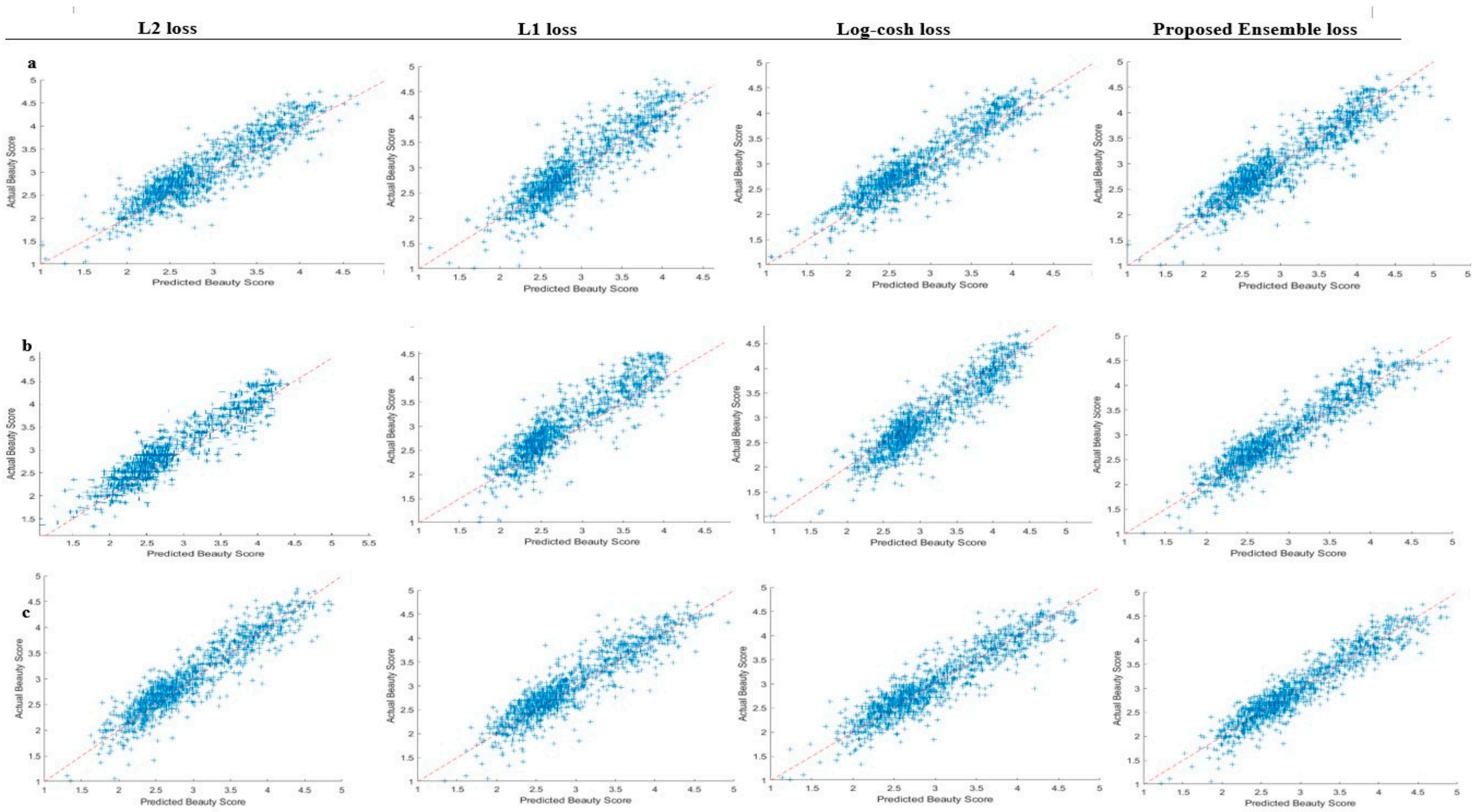

Figure 8. A scatter-plot representation of samples of predicted beauty scores against actual beauty scores on SCUT-FBP5500: (**a**) AlexNet, (**b**) VGG16-Net, and (**c**) FIAC-Net.

Table 7. Five-fold cross-validation of FBP, assuming different loss functions, utilizing three diverse CNNs on SCUT-FBP5500.

		AlexNet			**VGG16-Net**			**FIAC-NET**		
Loss	**Fold #**	**PC ↑**	**MAE ↓**	**RMSE ↓**	**PC ↑**	**MAE ↓**	**RMSE ↓**	**PC ↑**	**MAE ↓**	**RMSE ↓**
L2	Fold1	0.907345	0.240533	0.307498	0.923767	0.213235	0.272778	0.921218	0.208841	0.267048
	Fold2	0.907436	0.284240	0.353902	0.918951	0.240099	0.303619	0.923606	0.203690	0.263316
	Fold3	0.914046	0.292779	0.359794	0.929376	0.243384	0.300318	0.922841	0.215686	0.274935
	Fold4	0.910866	0.253034	0.313157	0.933455	0.234888	0.290986	0.930924	0.200899	0.258841
	Fold5	0.923891	0.206978	0.263612	0.932358	0.203312	0.260058	0.927243	0.211036	0.270286
	Average	**0.9127168**	**0.2555128**	**0.3195926**	**0.9275814**	**0.2269836**	**0.2855518**	**0.9251664**	**0.2080304**	**0.2668852**
L1	Fold1	0.897638	0.279448	0.351358	0.897066	0.248373	0.315218	0.907323	0.224207	0.295496
	Fold2	0.897693	0.233345	0.308535	0.903645	0.227187	0.304121	0.913023	0.207325	0.27959
	Fold3	0.898183	0.239692	0.313157	0.923759	0.223212	0.283766	0.901157	0.229243	0.301759
	Fold4	0.890245	0.243618	0.320583	0.899807	0.293189	0.368715	0.918057	0.212381	0.280385
	Fold5	0.888500	0.240513	0.315462	0.929065	0.247787	0.302772	0.922134	0.203054	0.267709
	Average	**0.8944518**	**0.2473232**	**0.321819**	**0.9106684**	**0.2479496**	**0.3149184**	**0.9123388**	**0.215242**	**0.284987**
Log-cosh	Fold1	0.906932	0.231047	0.296844	0.916811	0.221005	0.282706	0.916166	0.215914	0.279431
	Fold2	0.909971	0.258805	0.324433	0.912169	0.220701	0.283236	0.922908	0.207376	0.265653
	Fold3	0.908483	0.242075	0.305146	0.932647	0.225791	0.282773	0.928752	0.208598	0.266964
	Fold4	0.912021	0.291740	0.359610	0.916201	0.229308	0.297403	0.922272	0.208786	0.272133
	Fold5	0.921644	0.209866	0.268154	0.931253	0.229251	0.295928	0.929894	0.216112	0.269370
	Average	**0.9118102**	**0.2467066**	**0.3108374**	**0.9218162**	**0.2252112**	**0.2884092**	**0.9239984**	**0.2113572**	**0.2707102**
Proposed loss	Fold1	0.909468	0.234892	0.295199	0.931563	0.206476	0.272374	0.932550	0.195362	0.248635
	Fold2	0.910616	0.227180	0.297651	0.928253	0.239862	0.307400	0.922096	0.204956	0.272749
	Fold3	0.910364	0.227177	0.292339	0.931084	0.220507	0.279881	0.935712	0.199025	0.255806
	Fold4	0.912465	0.243282	0.312587	0.925647	0.225874	0.284645	0.925991	0.227766	0.287912
	Fold5	0.927530	0.291501	0.350512	0.938053	0.198880	0.251496	0.936200	0.186978	0.241975
	Average	**0.9140886**	**0.2448064**	**0.3096576**	**0.93092**	**0.2183198**	**0.2791592**	**0.9305098**	**0.2028174**	**0.2614154**

4.3.3. Performance Evaluation on MEBeauty

The anticipated presence of a substantial number of outliers in the MEBeauty dataset is attributed to the non-normal distribution of its data. Therefore, adopting CNNs with loss ensembles is necessary to obtain superior results over conventional regression-loss methods. Table 8 displays the results of five-fold cross-validation conducted on three pretrained CNNs, utilizing various loss functions for the MEBeauty dataset. It was revealed that L1 and L2 loss functions exhibit competitive results within AlexNet. However, for both VGG16 and FIAC-Net, L1 loss produces slightly lower PC values and higher error rates compared to L2 loss. In contrast, the Log-cosh loss function consistently yields improved results compared to both L1 and L2 loss functions.

Upon integration of the proposed ensemble loss function with AlexNet, the outcome consisted of a PC value of 0.8976712, MAE of 0.476394, and RMSE of 0.6097078. In contrast, VGG16-Net exhibited a slightly improved PC value of 0.907883. However, it also exhibited a slightly higher error rate, recording values of 0.512113 and 0.636318 in terms of MAE and RMSE, respectively, when compared to AlexNet.

Regarding FIAC-Net performance, it works well with a PC value of 0.925977, MAE of 0.426317, and RMSE of 0.536646. The reason behind the superiority of the fine-tuned FIAC-Net is that it was previously well-trained in [41] on more than 200,000 challenging facial images to efficiently classify the attractiveness of facial images. Meanwhile, AlexNet and VGG16-Net were pretrained on the vast ImageNet dataset for an object-classification task. Figure 9 illustrates the effectiveness of the proposed fine-tuned FIAC-Net with ensemble loss functions in predicting the beauty scores of MEBeauty-tested data. Meanwhile, Figure 10 presents a visual representation that facilitates a direct comparison between the estimated values and the true values of utilized CNNs that were implemented on the MEBeauty dataset.

The MEBeauty dataset presents a challenging and diverse collection of facial images captured under unconstrained conditions. It distinguishes itself through its larger scale in comparison to the SCUT-FBP dataset, which focuses exclusively on frontal images of Asian females, while remaining smaller than the SCUT-FBP5500 dataset, which encompasses frontal images of both Asian and Caucasian individuals. It is crucial to acknowledge that the size of a dataset alone does not guarantee improved model performance. Additional factors, such as gender, facial expressions, and poses of the subjects, significantly influence the model's efficacy. By accounting for these factors, researchers can obtain valuable insights into the intricate trade-off between image quality, dataset size, and their profound impact on the model's findings and outcomes. In this study, the proposed model demonstrated its effectiveness through efficient pretraining on the CelebA dataset for facial-attractiveness classification.

Table 8. Five-fold cross-validation of FBP, assuming different loss functions, using three diverse CNNs on MEBeauty.

		AlexNet			**VGG16-Net**			**FIAC-NET**		
Loss	**Fold #**	**PC ↑**	**MAE ↓**	**RMSE ↓**	**PC ↑**	**MAE ↓**	**RMSE ↓**	**PC ↑**	**MAE ↓**	**RMSE ↓**
L2	Fold1	0.874151	0.539646	0.679715	0.9149	0.657429	0.788001	0.908299	0.446347	0.569966
	Fold2	0.890606	0.483732	0.617635	0.89946	0.53349	0.671474	0.913656	0.440536	0.567868
	Fold3	0.885824	0.543017	0.688969	0.903028	0.494082	0.619434	0.919398	0.429767	0.5439
	Fold4	0.878471	0.50383	0.640299	0.901086	0.525557	0.658758	0.909376	0.469771	0.590058
	Fold5	0.878958	0.528132	0.687316	0.889669	0.477722	0.613175	0.907632	0.487113	0.614645
	Average	**0.881602**	**0.519671**	**0.662787**	**0.901629**	**0.537656**	**0.670168**	**0.911672**	**0.454707**	**0.577287**
L1	Fold1	0.876628	0.484963	0.651704	0.893344	0.588461	0.735082	0.898483	0.507801	0.640425
	Fold2	0.880223	0.499715	0.639342	0.908924	0.500534	0.714772	0.907686	0.505487	0.638085
	Fold3	0.888857	0.471078	0.619127	0.881992	0.555658	0.689619	0.899946	0.410486	0.549805
	Fold4	0.881179	0.578539	0.744373	0.884385	0.49881	0.634568	0.915355	0.458133	0.589622
	Fold5	0.87822	0.471296	0.649191	0.877244	0.514752	0.662075	0.908439	0.480193	0.612392
	Average	**0.881021**	**0.501118**	**0.660747**	**0.889178**	**0.531643**	**0.687223**	**0.905982**	**0.47242**	**0.606066**
Log-cosh	Fold1	0.888677	0.483999	0.620113	0.904497	0.593918	0.72576	0.916462	0.425651	0.535359
	Fold2	0.897537	0.469122	0.600731	0.907009	0.483545	0.606491	0.918844	0.454527	0.576362
	Fold3	0.885437	0.456751	0.580649	0.887036	0.466082	0.605073	0.915077	0.41715	0.525493
	Fold4	0.871824	0.514141	0.669348	0.920185	0.532153	0.662263	0.916048	0.438528	0.547494
	Fold5	0.880027	0.573935	0.740608	0.893962	0.502262	0.630416	0.920189	0.443971	0.570682
	Average	**0.8847**	**0.49959**	**0.64229**	**0.9025378**	**0.515592**	**0.6460006**	**0.917324**	**0.435965**	**0.551078**
Proposed loss	Fold1	0.898394	0.462036	0.606094	0.910804	0.486603	0.602119	0.92504	0.422594	0.5347
	Fold2	0.908921	0.451912	0.573237	0.919966	0.44634	0.560829	0.925186	0.422061	0.532776
	Fold3	0.906027	0.466733	0.58976	0.900498	0.542547	0.673791	0.927617	0.428647	0.533562
	Fold4	0.887227	0.504793	0.636339	0.910808	0.616375	0.74905	0.9295	0.413832	0.515335
	Fold5	0.887787	0.496496	0.643109	0.897337	0.468702	0.595801	0.922544	0.444453	0.566857
	Average	**0.8976712**	**0.476394**	**0.6097078**	**0.907883**	**0.512113**	**0.636318**	**0.925977**	**0.426317**	**0.536646**

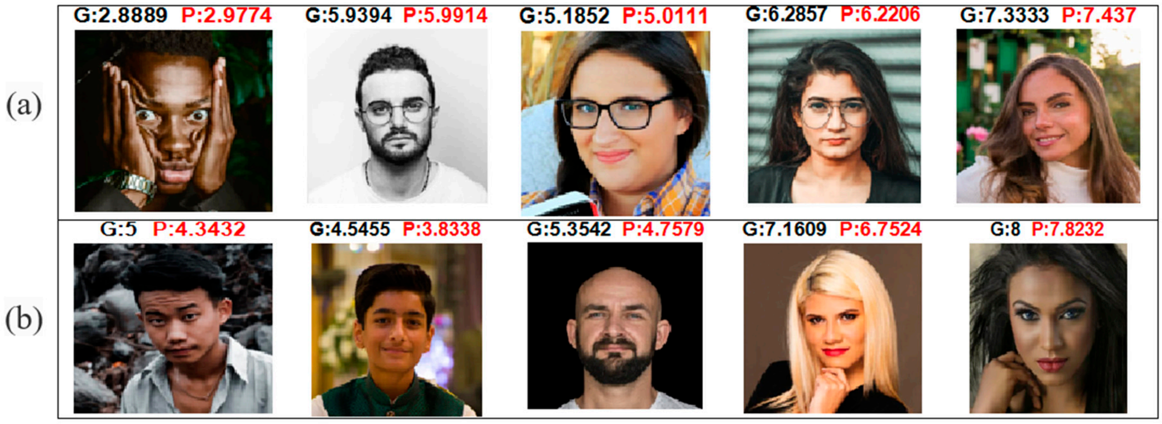

Figure 9. MEBeauty samples tested on the FIAC-Net. G, ground truth; P, model-predicted score. (**a**) Accurately predicted instance and (**b**) inaccurately predicted instances.

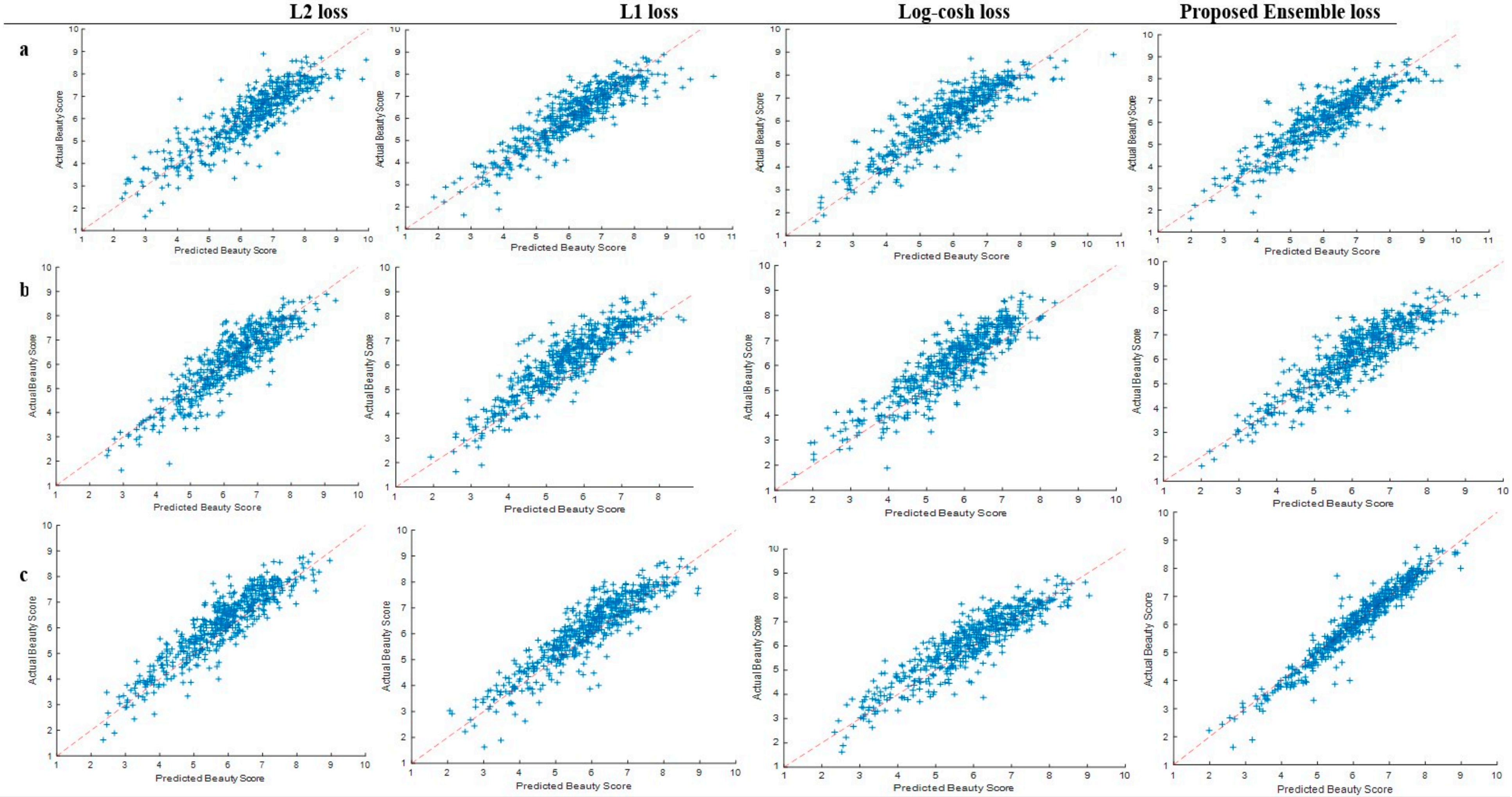

Figure 10. A scatter-plot representation of samples of predicted beauty scores against actual beauty scores on MEBeauty: (**a**) AlexNet, (**b**) VGG16-Net, and (**c**) FIAC-Net.

4.4. A Comparison to Other FBP-Related Work

We compared different FBP models based on three distinct datasets, namely the SCUT-FBP, SCUT-FBP5500, and MEBeauty datasets. The proposed method was compared with existing studies on these datasets, and the results are presented in Table 9. Our proposed FBP methodology outperformed the state-of-the-art techniques, as evidenced by the PC, RMSE, and MAE values. "N/A" is used to indicate that these studies do not provide the values of MAE and RMSE. While values in bold signify superior performance.

Table 9. Performance comparison with FBP state-of-the-arts.

Method	SCUT-FBP			SCUT-FBP 5500			ME Beauty		
	↑ PC	↓ MAE	↓ RMSE	↑ PC	↓ MAE	↓ RMSE	↑ PC	↓ MAE	↓ RMSE
VGG16 + Bayesian Ridge Regression [29]	0.857	0.2595	0.3397	N/A	N/A	N/A	N/A	N/A	N/A
ResNeXt-50-based R3CNN [22]	**0.95**	0.2314	0.2885	0.9055	0.2236	0.2954	N/A	N/A	N/A
HMTNet [32]	0.8977	N/A	N/A	0.8783	0.2501	0.3263	N/A	N/A	N/A
ResNet-18 based P-AaNet [48]	0.9103	0.2224	0.2816	0.9055	0.2236	0.2954	N/A	N/A	N/A
Cascade fine-tuned CNN [24]	0.88	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
REX-INCEP [35]	N/A	N/A	N/A	92.18	0.2052	0.2698	N/A	N/A	N/A
CNN-ER [49]	N/A	N/A	N/A	0.9250	**0.2009**	0.2650	N/A	N/A	N/A
VGGFace2+ Ensemble Stack [50]	0.8898	0.2409	0.3105	0.9112	0.2304	0.2951	N/A	N/A	N/A
Ensemble CNN [28]	0.8795	0.226	0.330	0.886	0.242	0.320	0.888	**0.365**	0.600
AlexNet + proposed ensemble loss	0.9038	0.2635	0.3482	0.9141	0.2449	0.3097	0.8977	0.4764	0.6097
VGG16-Net + proposed ensemble loss	0.9059	0.2229	0.2920	**0.9309**	0.2183	0.2792	0.9079	0.5121	0.6363
FIAC-Net + proposed ensemble loss	0.9101	**0.1859**	**0.2591**	0.9305	0.2028	**0.2614**	**0.9260**	0.4263	**0.5366**

In evaluating the SCUT-FBP dataset, the proposed fine-tuned FIAC-Net integrated with the proposed ensemble average lossl stands out as a top performer. It achieved a slightly higher PC of 0.9101, lower MAE of 0.1859, and lower RMSE of 0.2591 compared to most of the other investigated methods. Notably, the authors in [22] guided the regression task via pairwise ranking, which works well for small data, and it achieved a higher PC of 0.95, MAE of 0.2885, and RMSE of 0.2314, but our proposed model surpasses it in terms of lower error rates. Similarly, the proposed model exhibited an exceptional performance compared to other methods on the SCUT-FBP 5500 and MEBeaty datasets. It outperformed other compared approaches, showcasing superior performance in predicting facial-beauty scores. Our approach demonstrated competitive results on the SCUT-FBP5500 and MEBeauty datasets, achieving an MAE of 0.2028 and 0.4263, respectively. This performance is competitive to the FBP models presented in [28,49], which achieved MAEs of 0.2009 and 0.365, respectively. These results highlight the effectiveness of our approach in this particular context.

5. Conclusions

CNNs are a powerful method for making predictions, not only for classification problems but also for regression concerns. In regression, the focus is on understanding the relationship between continuous-number scores and data. The prediction network seeks to bring the estimated output closer to the actual scores by minimizing the average value of the loss function over the data related to the network weights. However, A CNN-regression-based model developed for FBP may face challenges if it relies solely on traditional loss functions. These challenges stem from biases naturally embedded in facial-beauty data, leading to an uneven distribution of data. For instance, preferences toward certain beauty scores, ethnicities, or age groups can introduce biases that impede the model's ability to generalize effectively across various situations. Accordingly, a new ensemble average loss function composed of three distinct regression-loss functions (L2 loss, L1 loss, and Log-cosh) was introduced by this work. and then integrated within various pretrained CNN architectures, namely AlexNet, VGG16-Net, and FIAC-Net. Its efficacy in predicting facial-image beauty scores was evaluated across three FBP benchmarks: SCUT-FBP, SCUT-FBP5500, and MEBeauty. It demonstrated its superiority when compared to the state-of-the-art. Additionally, our approach can potentially improve model performance by providing a significant correlation between machine- and human-predicted beauty scores and a low error rate. These findings highlight the effectiveness of the proposed ensemble cost function for regression tasks and suggest its potential use in improving CNN models.

Author Contributions: Conceptualization, J.N.S.; Methodology, J.N.S.; Writing—original draft, J.N.S.; Supervision, A.M.A. and D.A.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data related to this study can be accessed through the provided links: SCUT-FBP dataset: http://www.hcii-lab.net/data/SCUT-FBP/, SCUT-FBP5500 dataset: https://github.com/HCIILAB/SCUT-FBP5500-Database-Release, MEBeauty dataset: https://github.com/fbplab/MEBeauty-database, accessed on 5 June 2023.

Conflicts of Interest: The authors declare there is no conflict of interest.

References

1. Yang, C.-T.; Wang, Y.-C.; Lo, L.-J.; Chiang, W.-C.; Kuang, S.-K.; Lin, H.-H. Implementation of an Attention Mechanism Model for Facial Beauty Assessment Using Transfer Learning. *Diagnostics* **2023**, *13*, 1291. [CrossRef] [PubMed]
2. Gan, J.; Xie, X.; Zhai, Y.; He, G.; Mai, C.; Luo, H. Facial beauty prediction fusing transfer learning and broad learning system. *Soft Comput.* **2022**, *27*, 13391–13404. [CrossRef]
3. Chen, H.; Li, W.; Gao, X.; Xiao, B. Novel Multi-feature Fusion Facial Aesthetic Analysis Framework. *IEEE Trans. Big Data* **2023**, 1–18. [CrossRef]
4. Moridani, M.K.; Jamiee, N.; Saghafi, S. Human-like evaluation by facial attractiveness intelligent machine. *Int. J. Cogn. Comput. Eng.* **2023**, *4*, 160–169. [CrossRef]
5. Saeed, J.; Abdulazeez, A.M. Facial Beauty Prediction and Analysis Based on Deep Convolutional Neural Network: A Review. *J. Soft Comput. Data Min.* **2021**, *2*, 1–12. [CrossRef]
6. Liao, Y.; Deng, W. Deep Rank Learning for Facial Attractiveness. In Proceedings of the 2017 4th IAPR Asian Conference on Pattern Recognition (ACPR), Nanjing, China, 26–29 November 2017.
7. Belagiannis, V.; Rupprecht, C.; Carneiro, G.; Navab, N. Robust optimization for deep regression. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015.
8. Moshagen, T.; Adde, N.A.; Rajgopal, A.N. Finding hidden-feature depending laws inside a data set and classifying it using Neural Network. *arXiv* **2021**, arXiv:2101.10427.
9. Karal, O. Maximum likelihood optimal and robust support vector regression with lncosh loss function. *Neural Netw.* **2017**, *94*, 1–12. [CrossRef]
10. Huber, P.J. Robust estimation of a location parameter. In *Breakthroughs in Statistics: Methodology and Distribution*; Springer: Berline, Germany, 1992; pp. 492–518.
11. Black, M.J.; Rangarajan, A. On the unification of line processes, outlier rejection, and robust statistics with applications in early vision. *Int. J. Comput. Vis.* **1996**, *19*, 57–91. [CrossRef]
12. Merentitis, A.; Debes, C.; Heremans, R. Ensemble learning in hyperspectral image classification: Toward selecting a favorable bias-variance tradeoff. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 1089–1102. [CrossRef]
13. BenTaieb, A.; Kawahara, J.; Hamarneh, G. Multi-loss convolutional networks for gland analysis in microscopy. In Proceedings of the 2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI), Prague, Czech Republic, 13–16 April 2016.
14. Hajiabadi, H.; Molla-Aliod, D.; Monsefi, R. On extending neural networks with loss ensembles for text classification. *arXiv* **2017**, arXiv:1711.05170.
15. Xu, L.; Xiang, J. Comboloss for facial attractiveness analysis with squeeze-and-excitation networks. *arXiv* **2020**, arXiv:2010.10721.
16. Anderson, R.; Gema, A.P.; Isa, S.M. Facial attractiveness classification using deep learning. In Proceedings of the 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), Jakarta, Indonesia, 7–8 September 2018.
17. Dornaika, F.; Moujahid, A.; Wang, K.; Feng, X. Efficient deep discriminant embedding: Application to face beauty prediction and classification. *Eng. Appl. Artif. Intell.* **2020**, *95*, 103831. [CrossRef]
18. Xu, L.; Xiang, J.; Yuan, X. CRNet: Classification and regression neural network for facial beauty prediction. In *Pacific Rim Conference on Multimedia*; Springer: Berlin/Heidelberg, Germany, 2018.
19. Chen, F.; Zhang, D.; Wang, C.; Duan, X. Comparison and fusion of multiple types of features for image-based facial beauty prediction. In *Chinese Conference on Biometric Recognition*; Springer: Berlin/Heidelberg, Germany, 2017.
20. Dantcheva, A.; Dugelay, J.-L. Assessment of female facial beauty based on anthropometric, non-permanent and acquisition characteristics. *Multimed. Tools Appl.* **2014**, *74*, 11331–11355. [CrossRef]

21. Altwaijry, H.; Belongie, S. Relative ranking of facial attractiveness. In Proceedings of the 2013 IEEE Workshop on Applications of Computer Vision (WACV), Clearwater Beach, FL, USA, 15–17 January 2013.
22. 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**, *13*, 122–134. [CrossRef]
23. Hong, Y.-J.; Nam, G.P.; Choi, H.; Cho, J.; Kim, I.-J. A Novel Framework for Assessing Facial Attractiveness Based on Facial Proportions. *Symmetry* **2017**, *9*, 294. [CrossRef]
24. Xu, J.; Jin, L.; Liang, L.; Feng, Z.; Xie, D. A new humanlike facial attractiveness predictor with cascaded fine-tuning deep learning model. *arXiv* **2015**, arXiv:1511.02465.
25. Chen, F.; Xiao, X.; Zhang, D. Data-driven facial beauty analysis: Prediction, retrieval and manipulation. *IEEE Trans. Affect. Comput.* **2016**, *9*, 205–216. [CrossRef]
26. Zhai, Y.; Yu, C.; Qin, C.; Zhou, W.; Ke, Q.; Gan, J.; Labati, R.D.; Piuri, V.; Scotti, F. Facial Beauty Prediction via Local Feature Fusion and Broad Learning System. *IEEE Access* **2020**, *8*, 218444–218457. [CrossRef]
27. Iyer, T.J.; Nersisson, R.; Zhuang, Z.; Joseph Raj, A.N.; 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**, *2021*, 4423407. [CrossRef]
28. Saeed, J.N.; Abdulazeez, A.M.; Ibrahim, D.A. An Ensemble DCNNs-Based Regression Model for Automatic Facial Beauty Prediction and Analyzation. *Trait. Signal* **2023**, *40*, 55–63. [CrossRef]
29. Xu, L.; Xiang, J.; Yuan, X. Transferring rich deep features for facial beauty prediction. *arXiv* **2018**, arXiv:1803.07253.
30. Lebedeva, I.; Guo, Y.; Ying, F. Deep facial features for personalized attractiveness prediction. *SPIE* **2021**, *11878*, 72–80.
31. Gao, L.; Li, W.; Huang, Z.; Huang, D.; Wang, Y. Automatic facial attractiveness prediction by deep multi-task learning. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018.
32. Xu, L.; Fan, H.; Xiang, J. Hierarchical multi-task network for race, gender and facial attractiveness recognition. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019.
33. Lebedeva, I.; Guo, Y.; Ying, F. MEBeauty: A multi-ethnic facial beauty dataset in-the-wild. *Neural Comput. Appl.* **2022**, *34*, 14169–14183. [CrossRef]
34. Zhai, Y.; Huang, Y.; Xu, Y.; Gan, J.; Cao, H.; Deng, W.; Labati, R.D.; Piuri, V.; Scotti, F. Asian female facial beauty prediction using deep neural networks via transfer learning and multi-channel feature fusion. *IEEE Access* **2020**, *8*, 56892–56907. [CrossRef]
35. Dornaika, F.; Bougourzi, F.; Taleb-Ahmed, A.; Distante, C. Facial Beauty Prediction Using Hybrid CNN Architectures and Dynamic Robust Loss Function. In Proceedings of the International Conference on Pattern Recognition Workshop: Deep Learning for Visual Detection and Recognition, Montreal, QC, Canada, 21 August 2022.
36. Hajiabadi, H.; Monsefi, R.; Yazdi, H.S. relf: Robust regression extended with ensemble loss function. *Appl. Intell.* **2019**, *49*, 1437–1450. [CrossRef]
37. Muthukumar, V.; Narang, A.; Subramanian, V.; Belkin, M.; Hsu, D.; Sahai, A. Classification vs regression in overparameterized regimes: Does the loss function matter? *J. Mach. Learn. Res.* **2021**, *22*, 10104–10172.
38. Jadon, A.; Patil, A.; Jadon, S. A Comprehensive Survey of Regression Based Loss Functions for Time Series Forecasting. *arXiv* **2022**, arXiv:2211.02989.
39. Motepe, S.; Hasan, A.N.; Shongwe, T. Forecasting the Total South African Unplanned Capability Loss Factor Using an Ensemble of Deep Learning Techniques. *Energies* **2022**, *15*, 2546. [CrossRef]
40. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
41. Saeed, J.N.; Abdulazeez, A.M.; Ibrahim, D.A. FIAC-Net: Facial Image Attractiveness Classification Based on Light Deep Convolutional Neural Network. In Proceedings of the 2022 Second International Conference on Computer Science, Engineering and Applications (ICCSEA), Gunupur, India, 8 September 2022.
42. Liu, Z.; Luo, P.; Wang, X.; Tang, X. Deep learning face attributes in the wild. In Proceedings of the IEEE International conference on Computer Vision, Santiago, Chile, 7–13 December 2015.
43. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [CrossRef]
44. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
45. Xie, D.; Liang, L.; Jin, L.; Xu, J.; Li, M. Scut-fbp: A benchmark dataset for facial beauty perception. In Proceedings of the 2015 IEEE International Conference on Systems, Man, and Cybernetics, Kowloon Tong, Hong Kong, 9–12 October 2015.
46. Liang, L.; Lin, L.; Jin, L.; Xie, D.; Li, M. Scut-fbp5500: A diverse benchmark dataset for multi-paradigm facial beauty prediction. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018.
47. Dornaika, F.; Wang, K.; Arganda-Carreras, I.; Elorza, A.; Moujahid, A. Toward graph-based semi-supervised face beauty prediction. *Expert Syst. Appl.* **2020**, *142*, 112990. [CrossRef]
48. Lin, L.; Liang, L.; Jin, L.; Chen, W. Attribute-Aware Convolutional Neural Networks for Facial Beauty Prediction. In Proceedings of the IJCAI, Macao, China, 10–16 August 2019.

49. 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*, 108246. [CrossRef]
50. Vahdati, E.; Suen, C.Y. Female facial beauty analysis using transfer learning and stacking ensemble model. In Proceedings of the Image Analysis and Recognition: 16th International Conference, ICIAR 2019, Waterloo, ON, Canada, 27–29 August 2019; Proceedings, Part II 16; Springer: Berlin/Heidelberg, Germany, 2019.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.