engineering proceedings

Proceeding Paper

Facial Beauty Prediction using an Ensemble of Deep Convolutional Neural Networks †

Djamel Eddine Boukhari [1,*], Ali Chemsa [1], Abdelmalik Taleb-Ahmed [2], Riadh Ajgou [1] and Mohamed taher Bouzaher [3]

[1] LGEERE Laboratory Department of Electrical Engineering, University of El Oued, 39000 El-Oued, Algeria; d-technologie@univ-eloued.dz (A.C.); email1@email.com (R.A.)
[2] Univ. Polytechnique Hauts-de-France, Univ. Lille, CNRS, Centrale Lille Valenciennes, France; abdelmalik.taleb-ahmed@uphf.fr
[3] Scientific and Technical Research Centre for Arid Areas (CRSTRA), Biskra, Algeria; mohamedbouzaher2@gmail.com
* Correspondence: boukhari-djameleddine@univ-eloued.dz; Tel.: +213-660481464
† Presented at the 4th International Electronic Conference on Applied Sciences, 27 October–10 November 2023; Available online: https://asec2023.sciforum.net/.

Abstract: The topic of facial beauty analysis has emerged as a crucial and fascinating subject in human culture. With various applications and significant attention from researchers, recent studies have investigated the relationship between facial features and age, emotions, and other factors using multidisciplinary approaches. Facial beauty prediction is a significant visual recognition problem for the assessment of facial attractiveness, which is consistent with human perception. Overcoming the challenges associated with facial beauty prediction requires considerable effort due to the field's novelty and lack of resources. In this vein, a deep learning method has recently demonstrated remarkable abilities in feature representation and analysis. Accordingly, this paper contains main contributions propose an ensemble based on the pre-trained convolutional neural networks models to identify scores for facial beauty prediction. These ensembles are three separate deep convolutional neural networks, each with a unique structural representation built by previously trained models from Inceptionv3, Mobilenetv2 and a new simple network based on Convolutional Neural Networks (CNNs) for facial beauty prediction problem. According to the SCUT-FBP5500 benchmark dataset the model obtains 0.9350 Pearson Coefficient Experimental results demonstrated that using this ensemble of deep network leads to better predicting of facial beauty closer to human evaluation than conventional technology that spreads the facial beauty. Finally, potential research directions are suggested for future research on facial beauty prediction.

Keywords: deep learning; convolutional neural networks; facial beauty prediction; performance evaluation

Citation: Boukhari, D.E.; Chemsa, A.; Taleb-Ahmed, A.; Ajgou, R.; taher Bouzaher, M. Facial Beauty Prediction using an Ensemble of Deep Convolutional Neural Networks. *Eng. Proc.* **2023**, *52*, x. https://doi.org/10.3390/xxxxx

Academic Editor(s): Name

Published: date

1. Introduction

The human face holds a distinct importance in our social interactions, and the pursuit of beauty, particularly facial beauty, is an enduring and ubiquitous feature of human nature. In recent years, there has been a significant increase in the demand for aesthetic surgery, underscoring the importance of a nuanced understanding of beauty in medical settings [1]. Remarkably, the exploration of physical beauty in humans has a storied history dating back over 4000 years, demonstrating the enduring relevance of this topic [2]. The importance of physical beauty in the face has been studied for hundreds of years, and its influence on social decisions such as partner choices and hiring decisions is well documented [3]. The perception of facial attractiveness is considered a highly desirable physical trait, with philosophers, artists, and scientists all attempting to understand the secrets of beauty [4].

Facial beauty prediction is an emerging topic that is receiving increasing attention from researchers and users alike, particularly in the field of facial recognition and understanding [5]. Beauty is viewed as a form of information in computer-based face analysis, and it is linked to how people perceive attractiveness. In the field of psychology, several theories have been established on how people observe facial attractiveness. However, studying face attractiveness using computers is a relatively recent research area, with limited resources and few articles published on the subject. Several works have focused on analysing the irregular features of face attractiveness [6].

The analysis of facial attractiveness presents two main challenges. Firstly, the complexity of human perception and the wide variety of facial features make it difficult to build robust and effective models for evaluating beauty. Secondly, many face reference databases are primarily configured for face recognition problems and are not suitable for attractiveness prediction [7]. Therefore, most facial beauty studies focus on designing facial beauty descriptors [8].

In recent years, most facial beauty prediction research has been based on deep learning methods [9]. The development of deep learning architecture has been driven by the strength and adaptability of these algorithms, particularly convolutional neural networks (CNNs) [10]. These algorithms offer a novel perspective on the facial beauty prediction problem and have shown promising results for several computer vision applications such as face recognition, object identification, semantic segmentation, image classification, biomedical analysis, captioning, and biometrics [11]. DCNNs perform much better [12]. We create new ensembles models for face attractiveness evaluation. As a result, this study proposes ensembles are three separate deep convolutional neural networks for the facial beauty prediction (FBP) problem. The following are the present paper contributions:

- The investigation of the effectiveness of conventional transfer learning techniques on face beauty prediction.
- We provide an ensemble regression for face attractiveness evaluation using the projected scores of networks with three branches network trained InceptionV3, MobileNetV2 and new simple network based on Convolutional Neural Networks is proposed with loss functions.
- The efficiency of the suggested approach is demonstrated using the specialized FBP dataset, SCUT-FBP-5500. The efficiency of merging the assessments of several predictors in the proposed ensemble DCNNs regression model, which is considerably compatible with the ground truth of the dataset used, is demonstrated by the findings, which are encouraging. We have made our scripts and pre-processed pictures available to the general public at (https://github.com/DjameleddineBoukhari/ENCNN)

The structure of this paper is as follows: Some related researches on face attractiveness prediction are included in Section 2. Section 3 explains the selection process for the used architectures. On the SCUT-FBP5500 data set, Section 4 gives experimental findings and performance assessments.

2. Convolution Neural Networks Architecture for FBP

The initial concept of neural networks was extended by human nervous system, by mimicking the human nervous system. Scientists propose the concept of neural networks. Convolutional neural networks are further improvements over the neural network concept. The arrival of this model is good news for Auto-vision [13].

Numerous techniques based on deep neural networks (DNN) have been developed for FBP. One of the most popular CNN architectures, ResNet [27], has been utilized in a number of computer vision applications.

K. Cao et al. [5], used residual-in-residual (RIR) groups to build a deeper network. A combined spatial-wise and channel-wise attention mechanism is introduced for better feature comprehension. Authors presented their face beauty database SCUT-FBP5500 [15], with two evaluation protocols (5-fold cross validation 80–20% and 60–40% split). They tested three CNN architectures Alexnet [10], Resnet-18 [14] and ResneXt-50 [14].the results show better feature comprehension. In [16]. R3CNN architecture is proposed to integrate relative ranking into regression to improve the performance of FBP, and it can be flexibly implemented using existing CNNs as backbone network. This architecture provides better results than SCUT-FBP [17] and SCUT-FBP5500 [15] dataset.

3. Methodology

In order to facial beauty prediction, this work builds an ensemble of trained models. Mainly, our proposed approach EN-CNN architectures focus is on three pre-trained models and finally, the estimated of both are combined to create a final predicting facial beauty [25]. Transfer learning is used by the pre-trained models to reduce their weights so that they can perform a comparable regression task, for facial beauty an ensemble learning of trained models achieves greater performance. Consequently, in this study, we transfer the weights of the set of three potent pretrained CNN models. The next part presents the planned ensemble learning as well as the pre-trained Deep CNN models [26].

3.1. Pre-Trained InceptionV3

The Inception V3 [18] CNN was introduced by Google teams in [11]. Based on the Inception V1 model, Inception V3’s architecture was updated. The InceptionV3’s design incorporates numerous different kernel types at the same level. Instead of using a huge filter 7 × 7 and 5 × 5, the InceptionV3 uses a modest filter size 1 × 7 and 1 × 5. Furthermore, a bottleneck of 1 × 1 convolutions is used. Improved feature representation as a result. Beginning with the input data, three distinct convolutional layers with a 3 × 3 or 5 × 5 filter sizes are created by mapping parallel calculations. These layers’ output is combined into one layer, which is known as the output layer.

3.2. Pre-Trained MobileNetV2

A lightweight CNN model called MobileNet [19] is built on inverted residuals and a linear bottleneck, which provide quick connections between the thin layers. It is a low-latency model and uses a little amount of power therefore it is made to manage hardware resources that are limited. The MobileNet’s key benefit is the trade-off it makes between several elements including latency, accuracy, and resolution. In MobileNet, feature maps are generated using point-wise convolutional kernels and depth separable convolutional (DSC) kernels. In MobileNet [19], DSC first filters the input image’s spatial dimensions using depth-wise kennel 2-D filters. The depth-wise filter has a size of Dk × Dk × 1, which is significantly less than the size of the input images [24].

3.3. S-CNNs Network

In order to predict facial beauty, this work builds a simple CNNs. Mainly, our proposed approach S-CNNs architecture in constructed of several convolution layers and one fully connected layer at the end. Within each convolution layer, a 2D convolution is carried out, followed by ReLU activation. To demonstrate the effectiveness of the present S-CNNs, we control recent progress in neural architecture search to develop a new family of MixConv-based models. Our neural architecture are an ensemble of simple CNNs models, where the contribution is residing in the method of layer mixing with 3 × 3, 5 × 5, and 7 × 7 kernels size. The architecture is indicated in Figure 1.

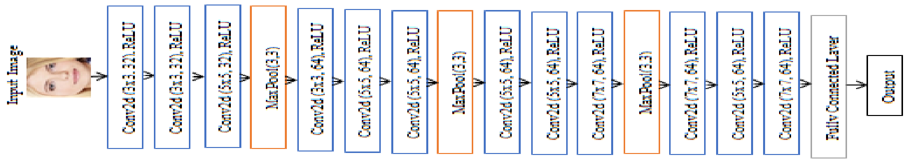

Figure 1. The architecture of S-CNNs network.

3.4. The Proposed EN-CNNs

The proposed ensemble of deep CNNs (EN-CNNs) architecture is following the three previously trained models (InceptionV3, MobileNetV2 and S-CNN). In the proposed EN-CNNs for the automated classification system, they act as basic classifiers of facial beauty. The details of the proposed EN-CNNs are as follows:

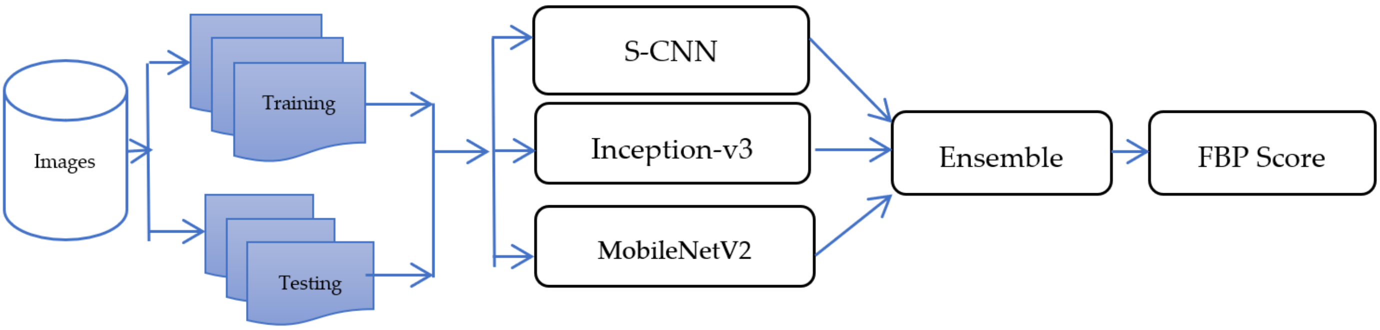

Figure 2. The proposed deep CNN ensemble networks (EN-CNNs).

4. Experiment

This section describes the experiments and assessment findings from the models (EN-CNN) used in this study. The SCUT-FBP5500 dataset [15] is used for network training. Our network is trained for 200 iteration with batch size of b = 25. The Adam optimizer updates the parameters with learning rate lr = 1e − 6. The selected loss function was MSE.

4.1. The SCUT-FBP5500 Dataset

The standard SCUT-FBP5500 dataset [15] is introduced in this study and it comprises 5,500 frontal face images at 350 × 350 resolutions with various attributes, including race (Asian/Caucasian), gender (female/male), and age (15–60) [20].

As shown in Figure 3 shown the ground truth rating for each face in the dataset is the average of all evaluations given on a scale from 1 to 5 by the 60 ratters. This enables the use of various computational models with various facial attractiveness prediction paradigms. The 2000 Asian females (AF), 2000 Asian men (AM), 750 Caucasian females (CF), and 750 Caucasian males (CM) are the four subsets of the SCUT-FBP5500 Dataset that may be separated according to race and gender [21].

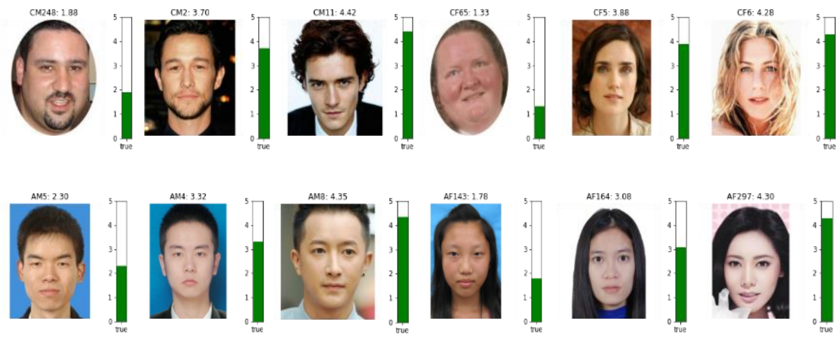

Figure 3. Images of various facial features and beauty ratings from the SCUT-FBP5500 benchmark dataset.

4.2. Results

Automatic face attractiveness estimate is achieved using an ensemble DCNNs regression. The experimental findings from the suggested EN-CNN are presented in this subsection. We carry out comparisons utilizing a range of techniques, including geometric feature-based and deep learning-based techniques, such as AlexNet, ResNet-18, ResNeXt-50, CNN–SCA, R3CNN and Semi-supervised etc. MAE, RMSE and PC are chosen as the metrics. Table 1 show the performance comparisons on the SCUT-FBP5500 dataset of facial beauty prediction using testify the EN-CNN capacity via comparison, which holds 80–20% splitting.

Table 1. Performance comparisons on the SCUT-FBP5500 dataset.

Methods	Pre-Training	MAE ↓	RMSE ↓	PC ↑
AlexNet [15]	ImageNet	0.2651	0.3481	0.8634
ResNet-18 [15]	ImageNet	0.2419	0.3166	0.8900
ResNeXt-50 [15]	ImageNet	0.2291	0.3017	0.8997
CNN–SCA [5]	ImageNet	0.2287	0.3014	0.9003
R3CNN [16]	ImageNet	0.2120	0.2800	0.9142
Semi-supervised [20]	VGGFace2	0.2210	0.2870	0.9113
CNN-ER [22]	VGGFace2	0.2009	0.2650	0.9250
NAS4FBP Net [23]	ImageNet	0.1939	0.2579	0.9275
EN-CNN Ours	ImageNet	**0.1933**	**0.2482**	**0.9350**

4.3. Discussion

Typically, the quantity of parameters imposes a limit on the performance improvement. The present model performs better than other models (AlexNet, ResNet-18, ResNeXt-50, CNN–SCA, R3CNN and Semi-supervised etc.). It uses an ensemble of deep CNNs (EN-CNNs) architecture the three models. Our EN-CNNs holds 26.99 M parameters, InceptionV3 with 22.85 M parameters model, MobileNetV2 with 2.91 M parameters model and Our network S-CNN holds 1.23 M parameters and 224.16 MFlops. CNN–SCA has 6.75 M parameters and 34.25 BFlops. ResNeXt-50 has 25.03 M parameters and 5.56 BFlops. AlexNet has 62.38 M parameters and 1. 5 BFlops. The comparison reveals that our network is better than the cited works. This tends to confirm that both the proposed EN-CNNs Network played a crucial role in outperforming the State-of-the-Art methods. In Figure 5, we visualize the comparisons of predicted scores.

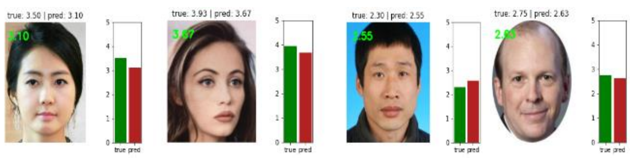

Figure 5. Comparisons of the ground-truth, and predicted scores given by EN-CNNs.

5. Conclusions

In this paper, an ensemble of deep CNNs for the facial beauty prediction is proposed. The study of power the standard transfer learning approaches on the facial beauty prediction problem, by combining the predicted scores of networks with three branches network (InceptionV3, MobileNetV2 and S-CNN) trained with loss functions, we present an ensemble regression for facial beauty estimation. Describe and optimize a set of hyperparameters for the new set of pre-trained models to classifier facial beauty. By utilizing an ensemble of all the previously mentioned transfer learning techniques, an ensemble (EN-CNN) was developed to predicting scores in facial beauty. The experimental findings show that our network can perform better than previous CNN baselines approaches. Experimental results showed that the proposed network achieved better performance as compared to several works available on the open literature. It improves the assessment's congruence with human judgment. As perspective, we propose to expand the scope of database and improve network using different architectures collected from Transfomer and ResNeSt.

Author Contributions:

Funding:

Institutional Review Board Statement:

Informed Consent Statement:

Data Availability Statement:

Acknowledgments This work was partially funded by Directorate General for Scientific Research and Technological Development (DGRSDT), Ministry of Higher Education and Scientific Research–Algeria.

Conflicts of Interest:

References

1. Zhang, D.; Chen, F.; Xu, Y. *Computer Models for Facial Beauty Analysis*; Springer International Publishing: Cham, Switzerland, 2016.
2. Fan, J.; Chau, K.; Wan, X.; Zhai, L.; Lau, E. Prediction of facial attractiveness from facial proportions. *Pattern Recognit.* **2012**, *45*, 2326–2334.
3. Helen, K.; Keith, O. Ranking facial attractiveness. *Eur. J. Orthod.* **2005**, *27*, 340–348.
4. 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.
5. Cao, K.; Choi, K.-N.; Jung, H.; Duan, L. Deep learning for facial beauty prediction. *Information* **2020**, *11*, 391.
6. Saeed, J.N.; 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.
7. Gray, D.; Yu, K.; Xu, W.; Gong, Y. Predicting facial beauty without landmarks. In Proceedings of the Computer Vision–ECCV 2010: 11th European Conference on Computer Vision, Heraklion, Crete, Greece, 5–11 September 2010; Springer: Berlin/Heidelberg, 2010.
8. Gan, J.; Xiang, L.; Zhai, Y.; Mai, C.; He, G.; Zeng, J.; Bai, Z.; Labati, R.D.; Piuri, V.; Scotti, F. 2M BeautyNet: Facial beauty prediction based on multi-task transfer learning. *IEEE Access* **2020**, *8*, 20245–20256.
9. Diao, H.; Hao, Y.; Xu, S.; Li, G. Implementation of Lightweight Convolutional Neural Networks via Layer-Wise Differentiable Compression. *Sensors* **2021**, *21*, 3464.
10. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, 1097–1105.
11. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
12. Szegedy C, Liu W, Jia Y. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and pattern Recognition, Boston, MA, USA, 7–12 June 2015.
13. Gan, J.; Jiang, K.; Tan, H.; He, G. Facial beauty prediction based on lighted deep convolution neural network with feature extraction strengthened. *Chin. J. Electron.* **2020**, *29*, 312–321.
14. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
15. 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; IEEE: Piscataway, NJ, USA, 2018.
16. 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.
17. 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; IEEE: Piscataway, NJ, USA, 2015.

18. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA, 26 June–1 July 2016.
19. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.
20. Dornaika, F.; Moujahid, A. Multi-View Graph Fusion for Semi-Supervised Learning: Application to Image-Based Face Beauty Prediction. *Algorithms* **2022**, *15*, 207.
21. Lebedeva, I.; Ying, F.; Guo, Y. Personalized facial beauty assessment: a meta-learning approach. *Vis. Comput.* **2022**, *39*, 1095–1107.
22. 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.
23. Zhang, P.; Liu, Y. NAS4FBP: Facial Beauty Prediction Based on Neural Architecture Search. In Proceedings of the International Conference on Artificial Neural Networks; Springer: Cham, Switzerland, 2022.
24. Srinivasu, P.N.; SivaSai, J.G.; Ijaz, M.F.; Bhoi, A.K.; Kim, W.; Kang, J.J. Classification of skin disease using deep learning neural networks with MobileNet V2 and LSTM. *Sensors* **2021**, *21*, 2852.
25. Lou, G.; Shi, H. Face image recognition based on convolutional neural network. *China Commun.* **2020**, *17*, 117–124.
26. Peng, S.; Huang, H.; Chen, W.; Zhang, L.; Fang, W. More trainable inception-ResNet for face recognition. *Neurocomputing* **2020**, *411*, 9–19.
27. Zhang, Hang, et al. Resnest: Split-attention networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022.