Facial Beauty Prediction using AlexNet Model

# Yuktha M

*Department of Information Science and Engineering*

*RNS Institute of Technology*

Bengaluru, India

yuktham753@gmail.com

# Pavithra Sonu J

*Department of Information Science and Engineering*

*RNS Institute of Technology*

Bengaluru, India

pavithrasonu0110@gmail.com

# Sushmitha G

*Department of Information Science and Engineering*

*RNS Institute of Technology*

Bengaluru, India

sushmithasushma63@gmail.com

# Dr. S Sathish Kumar

*Department of Information Science and Engineering*

*RNS Institute of Technology*

Bengaluru, India

s.sathishkumar@rnsit.ac.in

# Tanvi Prasad

*Department of Information Science and Engineering*

*RNS Institute of Technology*

Bengaluru, India

tanviprasad029@gmail.com

***Abstract*—*Facial beauty prediction (FBP) is a rising topic in the area of evaluating attractiveness, with the goal of making assessments which are more human-like. Since FBP is a regression problem, there are many data driven ways to determine the relationships among face traits and beauty ratings. Deep learning approaches have shown their incredible potential for feature representations and analysis of the data. Convolutional Neural Network (CNN) has demonstrated extremely outstanding performance in facial identification and understandings, and have been proven to be an effective way for exploring facial features. Recently, a lot of well-designed networks with efficient topologies have been researched for improved representation of ability. However, these models focus more on the efficient block but do not create an efficient data transmission route, resulting in a feature representation capacity which is sub-optimal. In this paper, we have taken the Alex Net model into consideration which is a deep convolutional network which has been created to consider a large range of colorful images (224x224x3). It has a range of 62 million parameters which are trainable. For better performance, an intricate network architecture for the FBP issue is given in this research.***

***Key Words*—Facial Attractiveness Prediction, CNN, AlexNet’s Model, SCUT-FBP5500 Dataset.**

1. ***INTRODUCTION***

Face Beauty analysis possess a wide range of applications in various domains. Facial attractiveness analysis is employed in a variety of sectors, including facial picture beautify APPs, plastic surgery, and face based pose analysis. Billions of photographs have been acquired and published to public networks and online platforms in the mobile computing era on every day basis, development of better image processing and analysis technology have been necessitated. Computer and data concentrated approaches for tackling the problems, inclusive of face identification, face expression recognition, facial beauty analysis, and so on, have been deviced in the recent days. Facial characteristics have such a substantial part in the assessment of FBP difficulties. The features are then inspected and compiled for aggregate analysis after they have been acquired. Existing are multiple data specific strategies for FB prediction with finger or ease learn descriptors to pick a better portrayal of face features. These models use sophisticated predictors that also are trained in an useable format to perform assessments with extracted features.



**Figure 1.** Face beauty has been predicted as an example. There are four varieties of face photographs for the sake of prediction, in descending order: Asian male and femal (AM and Af respectively), Caucasian origin Female and Male (CF and CM respectively). The ranks gradually fall from left to right.

As per the American Society for Aesthetic Cosmetic Procedures, over $12 money was spent on over 11 million surgical and non - surgical beauty treatments in the year 2013.Different facial attractiveness related applications have kept popping up in android and IOS APP stores, drawing a large attention. Data driven facial attractiveness analysis seeks to develop computer simulations of face enhancement and provide solutions to real world applications using perfect example. Researchers from a multitude of disciplines, including psychiatry, evolutionary theory, plastic surgery, computer technology, and others, have conducted studies of facial attractiveness.

1. ***RELATED WORK***

Face beauty estimate using images is a novel problem in computer vision. FBP was originally treated as a regression task in a database in 2015 [10]. Hand-crafted approaches [11–18] or pattern recognition [12,19–21] are employed in the survey to model and assess facial beauty. Geometry-based and appearance-based methods are two types of hand-crafted methods [16]. P. Aarabi et al. deviced an autonomous facial beauty assessment model based on the correlations between facial features (face appearance, eyes, brows, and mouth) and used the KNN algorithm to develop a beauty bridging in [15]. D. Zhang et al. analysed thousands of female’s and male’s images to assign them to a subspace of human face shape. Using a curve fitting invariant shape distance metric, they then employed a countable method to examine the presence of geometrical face attributes on person's face beauty. H. Yan introduced a new Cost Sensitive Ordinal Regression to assess the significance of trials in various classes in [16]. They used their Cost Sensitive Ordinal Regression on four different kinds of attributes: high/low intensity, Local Binary Pattern, Scale Invariant Feature Transform, and LE [24]. L. Liang et al. [12] made use of shallow predictors such as linear regression (LR), Gaussian-regression, and support-vector specific regression with feature extraction (SVR).

Deep learning models have at present been frequently utilised to test the attractiveness of faces in recent years. L. Liang et al. used two evaluation methodologies to test their facial beauty database (SCUT-FBP5500) in [12]. AlexNet [25], Resnet-18 [26], and ResneXt-50 [27] were the three CNN architectures they tested. The ResneXt-50 design stood up in comparison to the other two deep architectures (Alexnet and Resnet-18), according to their findings. Furthermore, when using alternative shal- low regressors, the deep architectures outperformed the hand-crafted features. To provide superior gradient transmission flow, K. Cao et al. used a residual-in -residual block to establish a deep network with multiple level skip connections. They also looked for the inherent association between feature maps using both channel specific and space specific focused processes. SCUT -FBP5500 [12] database was used to test their method, and it performed well. L. Lin et al. propose an R3CNN model with only two primary features in [21]. The initial component of it is a regression specific component, which consists of 2 similar regression specific subnets that bridge each facial picture to beauty rating in a consistent manner. The next component is a scoring feature that use the Siamese’s model to develop a couple wise ranks to direct the beauty prediction. On SCUT FBP [10] and SCUT FBP5500 [12] database, their architecture yielded good results. Adding to supervised-learning, semi supervised learning has shown to be effective in estimating facial beauty [28, 29]. F. Dornaika et al. introduced a multiple layered locally specific discriminative embedding model using feature identification as the primary stage in [29]. The most significant and distinguished aspects of the inputted face’s image or face’s descriptor are captured via feature selection.

1. ***PROPOSED SYSTEM***
2. *Dataset*

A lot of research in the area of facial detection and face beauty prediction is done with the help of SCUT-FBP5500 as benchmark for these calculations. This is due to the presence of diversity in the images of both Male and female genders. To capture the real life facial beauty the range of age is also extensive as images of people belonging from age group of 15 years to 60 years is present. This dataset consists of 2000 number of females from Asia images, 2000 number males from Asia images, 750 number of Caucasian origin Females and 750 photos of Caucasian origin Males , mapped into a resolution of 350 × 350.



**Figure 2.** Pipeline corresponding to the proposed method. The face is identified and then given as input to CNNs; concatenation of the feature maps of both convolution4 1 along with convolution5 1 layer and flattening these into a attribute vector for Bayesian-ridge regression as input.

1. *Architecture*

The proposed method Alex net method is used to analyze facial beauty. AlexNet is a CNN algorithm with a bit more depth. Alex Net was a DN network built by Alex Krizhevsky and coworkers in the year of 2012. It took the lead in the Image Net LSVRC-2010 competition, where it was used to analyze photographs. It also worked with a lot of GPUs.



**Figure 3.** AlexNet Model’s Architecture

Alex Net is a convolutional deep network designed to accommodate huge, multicolored photos (224x224x3). It consists of a total of 62.3 million learnable parameters.

**Alex Net model’s 11 layers are as follows:**

* Layer C1: First Layer of Convolution (96, 11×11)
* Layer S2: First Layer of Maximum Pooling (3×3)
* Layer C3: Second Layer of Convolution (256, 5×5)
* Layer S4: Second Layer of Maximum Pooling (3×3)
* Layer C5: Third Layer of Convolution (384, 3×3)
* Layer C6: Fourth Layer of Convolution (384, 3×3)
* Layer C7: Fifth Layer of Convolution (256, 3×3)
* Layer S8: Third Layer of Maximum Pooling (3×3)
* Layer F9: Full Connected Layer (4096)
* Layer F10: Full Connected Layer (4096)
* Layer F11: Full Connected Layer (1000)

**C1:- First Layer Of Convolution**

The Convolution Layer which is the AlexNet's first layer looks at the 2242243 image tensor as input. In a convolution process, it is using 96 (1111) kernels with a padding of two and stride of four . This produces a 555596 output tensor, which is then given to its subsequent layer with the of use a ReLu function. The layer had 34,944 parameters which are trainable.

**S2:- First Layer of Maximum Pooling**

The 2nd tier of Alex Net, the max pooling tier, receives layer C1 as an output and tensor 555596 as an input data. The kernel’s implementation of a zero-padded sub validation with two strides. The result was an output tensor 272796, which was implemented in the following third layer.

**C3:- Second Layer Of Convolution**

Third layer succeeding the second Alex Net's model is a convolution layer uses the outcome of tier S2, which is a 272796 tensor, as  a new input. In the convolution’s process, the process used 256 (55) number of kernels consisting two padding and one stride. This generated a 2727256 numbered tensor output, then fed through a ReLu activation procedure prior sent to following tier. There was a total of 614,656  parameters that were trainable in the tier, for an another total of 649,600 parameters which are trainable.

**S4:- Second Layer of Maximum Pooling**

The 4th layer coming after the third of Alex Net is a maximum pooling tier that acquired input from the previous C3 layer, a 2727256 tensor. It completed a sub sampling procedure with a zero-padded and with the help of a (33) kernel with a two stride, equivalent to S2 and this generated a 1313256 tensor output, which has been sent to the C5 using a ReLu’s activation procedure.

**C5:- Third Layer Of Convolution**

The Alex Net model's 5th layer is a convolution layer in which the output of layer S4 is accepted as an input, a 1313256 tensor output. In such a operation, it utilized (33) kernels in the range of 384 with a padding and stride equivalent of one. This resulted in a 1313384 output tensor, which was eventually delivered to the next layer through a ReLu activation. There have been 885,120 trainable parameters in the layer, for such a total of 1,534,720 trainable parameters.

**C6:- Fourth Layer Of Convolution**

The Alex Net model's 6th layer is a layer of convolution that accepted the output of layer C5, which is a 1313384 tensor. It applied the same convolution process as layer C5, and the output vector was identical. The result was also treated to a ReLu’s activation procedure. There are 1,327,488 parameters which are trainable in this layer, raising the summed up quantity of trained parameters to 2,862,208.

**C7:- Fifth Layer Of Convolution**

Alex Net model's 7th layer is a convolution layer which acquired a 1313384 tensor as input data from C6. In a combination procedure along with a padding and stride in the range of 1, it required 256 (33) kernels. A 1313256 output tensor was indeed the outcome. The output too was exposed to a ReLu’s activation procedure. There have been 884,992 trainable parameters in this tier, for a maximum of 3,747,200 parameters which are trainable.

**S8:- Third Layer of Maximum Pooling**

Alex Net model's 8th layer is a maximum pooling layer which takes a 1313256 tensor from C7 as the new input. It uses a thirty-three window region with just stride of two to execute a sub sampling procedure with a zero padded. This resulting in a 6628 tensor output, that is then passed through a ReLu’s procedure before being given to the next tier.

**F9:- First Full Connected Layer**

Alex Net model’s 9th level/layer is a full linked level that takes a thin 66248 tensor as input from layer S8. It utilizes a weighted sum procedure along with non-linear function. This ultimately generates a tensor output (40961), then passed through a ReLu’s procedure prior sending to the following level. The layer included 37,752,832 parameters which are trainable, making a sum of trainable parameters around 41,500,032 until this layer.

**F10:- Second Full Connected Layer**

10th layer is a Alex Net model’s second full connected tier that took a 40961 input tensor data from layer F9. It executed very similar kind as tier F9 along with generating similar tensor as output, which is then transferred to the following layer after passing through a ReLu’s activation procedure. There were 16,781,312 trainable parameters in this layer, maximizing the total sum of parameters which are trainable to 58,281,344.

**F11:- Third Full Connected Layer**

The F10 layer supplied a (40961) tensor as input to the network's 11th and the last layer, which is also a full connected layer. It yielded a 10001 output tensor, which was ultimately transmitted through a SoftMax’s activation procedure, just like layers F9 as well as F10. The layer had such a maximum of 4,097,000 parameters which are trainable, for a grand of 62,378,344 parameters which are trainable. The outcome of SoftMax’s activation procedure included the network predictions.

1. ***RESULT***

SCUT FBP5500 dataset have been used to train the proposed system. This model calculates the score of the face taking the various facial features into account to predict the beauty score. Besides the beauty score, every feature present on the face is taken into account and it's deviation is calculated. Along with that cosmetics suggestions are proposed.



**Figure 4.** Top models beautification scores

In the figure 4, as a reference, we can see the score given to top models in cosmetic industry based on their facial features and traits by plotting a graph of those faces.



**Figure 5.** The varied expressions and neutralized mode means of face features considered for a total of thirty-six subjects.

In Fig. 5, we can see the top 5 traits which are: right-most eye's placing, chin's placing, lip placing, nostrils proportion and position, and left-most eye's placing. The worst five features are the distance measured between right-most ear and lips, nostrils width, the length between the left-most eye and the nostril, the width of the forehead, and the left-most eye’s width.

1. ***CONCLUSION***

The training model which has been proposed by us uses the Alex net model. It is a face beauty score predictor that uses the Deep CNN model and is used for calculating the deviation based on the facial features. The deviation scores calculated after evaluating the facial features are used as corrections in the plastic surgery and cosmetic industry to beautify one's face. With the onset of digital media, it has also sought the attention of scientists and physicians along with the artists.

***REFERENCES***

1. K. Dion, E. Berscheid, E. Walster, What is beautiful is good, J. Pers. Soc. Psychol. 24 (3) (1972) 285.
2. J. Gan, L. Xiang, Y. Zhai, C. Mai, G. He, J. Zeng, Z. Bai, R.D. Labati, V. Piuri, F. Scotti, 2M BeautyNet: Facial beauty prediction based on multi-task transfer learning, IEEE Access 8 (2020) 20245–20256.
3. Y. Eisenthal, G. Dror, E. Ruppin, Facial attractiveness: Beauty and the machine, Neural Comput. 18 (1) (2006) 119–142.
4. X. Liu, T. Li, H. Peng, I.C. Ouyang, T. Kim, R. Wang, Understanding beauty via deep facial features, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPRW, 2019, pp. 246–256.
5. T. Alashkar, S. Jiang, Y. Fu, Rule-based facial makeup recommendation system, in: 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition, FG 2017, IEEE, 2017, pp. 325–330.
6. A. Laurentini, A. Bottino, Computer analysis of face beauty: A survey, Comput. Vis. Image Underst. 125 (2014) 184–199.
7. L. Liang, L. Jin, X. Li, Facial skin beautification using adaptive region-aware masks, IEEE Trans. Cybern. 44 (12) (2014) 2600–2612.
8. L. Xu, H. Fan, J. Xiang, Hierarchical multi-task network for race, gender and facial attractiveness recognition, in: 2019 IEEE International Conference on Image Processing, ICIP, IEEE, 2019, pp. 3861–3865.
9. O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, Imagenet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (3) (2015) 211–252.
10. D. Xie, L. Liang, L. Jin, J. Xu, M. Li, SCUT-FBP: A benchmark dataset for facial beauty perception, in: 2015 IEEE International Conference on Systems, Man, and Cybernetics, IEEE, 2015, pp. 1821–1826.
11. L. Xu, J. Xiang, X. Yuan, Transferring rich deep features for facial beauty prediction, 2018, arXiv preprint arXiv:1803.07253.
12. L. Liang, L. Lin, L. Jin, D. Xie, M. Li, SCUT-FBP5500: A diverse bench-mark dataset for multi-paradigm facial beauty prediction, in: 2018 24th International Conference on Pattern Recognition, ICPR, 2018.
13. D. Gray, K. Yu, W. Xu, Y. Gong, Predicting facial beauty without landmarks, in: European Conference on Computer Vision, Springer, 2010, pp. 434–447.
14. D. Zhang, Q. Zhao, F. Chen, Quantitative analysis of human facial beauty using geometric features, Pattern Recognit. 44 (4) (2011) 940–950.
15. P. Aarabi, D. Hughes, K. Mohajer, M. Emami, The automatic measurement of facial beauty, in: 2001 IEEE International Conference on Systems, Man and Cybernetics. e-Systems and e-Man for Cybernetics in Cyberspace (Cat. No. 01CH37236), vol. 4, IEEE, 2001, pp. 2644–2647.
16. H. Yan, Cost-sensitive ordinal regression for fully automatic facial beauty assessment, Neurocomputing 129 (2014) 334–342.
17. W.-C. Chiang, H.-H. Lin, C.-S. Huang, L.-J. Lo, S.-Y. Wan, The cluster assessment of facial attractiveness using fuzzy neural network classifier based on 3D moiré features, Pattern Recognit. 47 (3) (2014) 1249–1260.
18. J. Fan, K.P. Chau, X. Wan, L. Zhai, E. Lau, Prediction of facial attractiveness from facial proportions, Pattern Recognit. 45 (6) (2012) 2326–2334.
19. K. Cao, K.-n. Choi, H. Jung, L. Duan, Deep learning for facial beauty

prediction, Information 11 (8) (2020) 391.

1. L. Lin, L. Liang, L. Jin, W. Chen, Attribute-aware convolutional neural

networks for facial beauty prediction, in: IJCAI, 2019, pp. 847–853.

1. L. Lin, L. Liang, L. Jin, Regression guided by relative ranking using convolutional neural network (R3CNN) for facial beauty prediction, IEEE Trans. Affect. Comput. (2019).
2. T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns: Application to face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 28 (12) (2006) 2037–2041.
3. D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vis. 60 (2) (2004) 91–110.
4. Z. Cao, Q. Yin, X. Tang, J. Sun, Face recognition with learning-based

descriptor, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, 2010, pp. 2707–2714.

1. A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with

deep convolutional neural networks, in: Advances in Neural Information Processing Systems, 2012, pp. 1097–1105.

1. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
2. S. Xie, R. Girshick, P. Dollár, Z. Tu, K. He, Aggregated residual transformations for deep neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1492–1500.
3. F. Dornaika, K. Wang, I. Arganda-Carreras, A. Elorza, A. Moujahid, Toward graph-based semi-supervised face beauty prediction, Expert Syst. Appl. 142 (2020) 112990.
4. F. Dornaika, A. Moujahid, K. Wang, X. Feng, Efficient deep discriminant embedding: Application to face beauty prediction and classification, Eng. Appl. Artif. Intell. 95 (2020) 103831.
5. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking

the inception architecture for computer vision, in: Proceedings of the

IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp.2818–2826.