Deep Face Beautification

Jianshu Li [1,5]
jianshu@u.nus.edu

Chao Xiong [2]
superx21@gmail.com

Luoqi Liu [4]
llq667@gmail.com

Xiangbo Shu [3,4]
shuxb104@gmail.com

Shuicheng Yan [4]
eleyans@nus.edu.sg

[1] School of Computing, National University of Singapore, Singapore
[2] Department of Electrical and Electronic Engineering, Imperial College London, U.K.
[3] School of Computer Science and Engineering, Nanjing University of Science and Technology, P. R. China
[4] Department of Electrical and Computer Engineering, National University of Singapore, Singapore
[5] SAP Research and Innovation Singapore, SAP Asia Pte Ltd

ABSTRACT

The beautification of human photos usually requires professional editing softwares, which are difficult for most users. In this technical demonstration, we propose a deep face beautification framework, which is able to automatically modify the geometrical structure of a face so as to boost the attractiveness. A learning based approach is adopted to capture the underlying relations between the facial shape and the attractiveness via training the Deep Beauty Predictor (DBP). Relying on the pre-trained DBP, we construct the BeAuty SHaper (BASH) to infer the "flows" of landmarks towards the maximal aesthetic level. BASH modifies the facial landmarks with the direct guidance of the beauty score estimated by DBP.

Categories and Subject Descriptors

J.5 [**Arts and Humanities**]: Arts, fine and performing

Keywords

Facial beautification; Face morphing

1. INTRODUCTION

The pursuit of beauty roots deeply in the intrinsic spirit of human beings and people incline to show their most attractive side to others by retouching the portraits. However, the aesthetic enhancement of photos usually has high requirements on the skills of using some professional editing softwares, such as Photoshop. Such barrier makes it difficult for most users to beautify photos by themselves. Therefore, automatic beautification of human face photos becomes a topic of great importance, which is demonstrated in this work. Despite the importance and great commercial potential, prior efforts on the statistical study of beauty are limited. Most previous works [4, 3] focus on the analysis of facial attractiveness. The work in [2] enhances facial attractiveness, but the score predictor is built on a simple linear regressor, whose discriminative capability is limited in capturing the complex underlying rules of aesthetic evaluation.

The recent success of deep learning has stimulated many applications in various research fields [1]. Deep neural network is able to learn a highly abstract and nonlinear feature representation with regard to the given objective, which makes it a suitable tool to capture the implicit link between facial geometry structure and attractiveness. In this demo, we use deep neural networks to automatically beautify facial images. We first learn the relation between facial geometrical structure and beauty level. Then with a photo of a human face, we can predict the beauty score, based on which the structure of facial landmarks is reshaped towards the maximal level of beauty. The face image is then warped according to the transformed landmarks to produce the beautified face.

In our demo, we process the user's input face image and display the result in real time. The interface is shown in Figure 1.

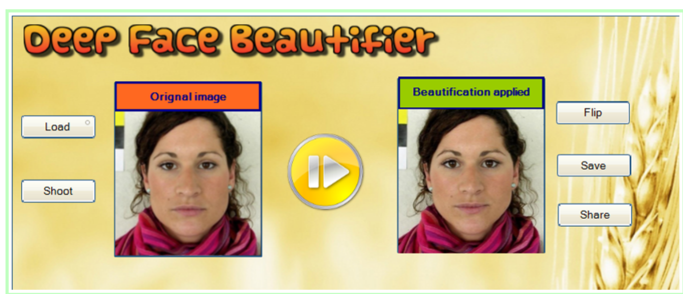

Figure 1: User interface of our demo

2. TECHNOLOGY

2.1 Training Dataset

Several datasets have been released for the studies on human attractiveness in literature. The Multi-Modality Beauty (M^2B) dataset in [4] is used to train our model. M^2B database includes facial, dressing images and audio files of 1,240 females, and provides corresponding attractiveness score for each modality. The scores are in the range of 1 to 10, where 1 is the lowest beauty level and 10 is the highest. During the training process of our model, only information in terms of the face images is employed.

MM'15, October 26–30, 2015, Brisbane, Australia.
ACM978-1-4503-3459-4/15/10.
DOI: http://dx.doi.org/10.1145/2733373.2807966.

2.2 Deep Beauty Predictor

A fully-connected deep neural network is adopted to learn the beauty score prediction, which is called Deep Beauty Predictor (DBP). The input feature vector of DBP is the concatenation of the facial landmarks' coordinates. Assume we are given n training samples $\{\boldsymbol{X}_i, i = 1, 2, ..., n\}$ with corresponding beauty scores $\{s_i, i = 1, 2, ..., n\}$. The DBP network aims at solving a regression problem with regard to the following objective

$$\mathcal{L} = \sum_{i=1}^{n} (\mathcal{F}(\boldsymbol{X}_i) - s_i)^2 + \alpha ||\boldsymbol{W}||^2, \tag{1}$$

where $\mathcal{F}$ denotes the feature mapping of the network, and $||\cdot||^2$ is the ℓ-2 regularizer on the network weights to reduce over-fitting. The objective can be optimized with standard Stochastic Gradient Descent (SGD).

In this demo, DBP follows a four-layered structure, and the numbers of hidden units are 800, 800, 300 and 1 respectively. Each of the first three layers is followed by ReLU as the activation function. The output of the last layer is used to compute the regression cost as in Eqn. (1), and thus the dimension should be 1. To reduce the impact of over-fitting, drop-out is adopted after each of the first three layers, and the drop-out rate is set as 0.3. The weight decay parameter α in Eqn. (1) is set as 0.1. As for the training of DBP, the initial learning rate is set to 0.001, and decreased by 50% every 2,000 iterations. The maximal iteration number is set as 20,000.

2.3 BeAuty SHaper with Deep Input Gradient

BeAuty SHaper (BASH) focuses on inferring the "flows" of landmark points such that the resulting beauty score of the transformed landmark locations is maximized. The flow is defined as a displacement vector, denoted as $\Delta\boldsymbol{T}$, from the original locations. The flows of landmarks define the characteristics of the reshaping motion. We inherit the pre-trained DBP in the BASH module, and utilize DBP to infer the landmark flows which can maximize the corresponding beauty score.

Let us denote the input landmark feature of the testing image i as $\boldsymbol{X}_i$, and the estimated beauty score of DBP can be formulated as

$$\tilde{s}_i = \mathcal{F}(\boldsymbol{X}_i + \Delta\boldsymbol{T}_i), \tag{2}$$

where $\mathcal{F}(\cdot)$ is the same mapping as described above and $\tilde{s}_i$ is the beauty score of the transformed landmark features. The objective is defined with regard to $\Delta\boldsymbol{T}_i$, and penalizes on the gap between the current beauty score and the maximal beauty score. In particular, we define a cost as follows,

$$\mathcal{L}_{beauty} = |\tilde{s}_i - M| \tag{3}$$

where M is the maximal beauty score, which is 10 on the currently used training dataset.

Based on Eqn. (2) and Eqn. (3), it can be observed that the gradient with regard to $\Delta\boldsymbol{T}_i$ actually equals the gradient in terms of the input vectors for DBP. In this work, the displacement vector is inferred via computing its gradient, referred to as the deep input gradient. By minimizing the cost in Eqn. (3), the network automatically moves the key landmarks towards the locations with the highest rating of beauty.

To retain the semantic relations between facial landmarks, we include a constraint defined with the aid of Probabilistic Principal Component Analysis (PPCA). We train a PPCA model on the high-score face images to capture the underlying distribution of the desired facial landmark structure. Accordingly, we penalize the negative log-likelihood of transformed landmark positions by introducing the following constraint:

$$\mathcal{L}_{ppca} = -ln(Pr(\boldsymbol{X}_i + \Delta\boldsymbol{T}_i | \boldsymbol{\mu}, \boldsymbol{P}\boldsymbol{P}^T + \sigma^2\boldsymbol{I})). \tag{4}$$

By minimizing $\mathcal{L}_{ppca}$, the transformed shape model is enforced to follow the same distribution of the targeted beauty group of high attractiveness. The resulting model, therefore, avoids the violation of relative semantic positions among facial landmarks.

In summary, the objective of the BASH module integrates the aforementioned two costs as

$$\mathcal{L} = \mathcal{L}_{beauty} + \beta_1 \cdot \mathcal{L}_{ppca}. \tag{5}$$

The corresponding gradient with regard to $\Delta\boldsymbol{T}_i$ is computed as follows,

$$\begin{aligned} \frac{\partial\mathcal{L}}{\partial\Delta T_i} = \quad & \frac{\partial\mathcal{L}_{beauty}}{\partial\Delta\boldsymbol{T}_i} + \\ & 2\beta_1 \left[(\boldsymbol{P}\boldsymbol{P}^T + \sigma^2\boldsymbol{I})^{-1} \cdot (\boldsymbol{X}_i + \Delta\boldsymbol{T}_i - \boldsymbol{\mu}) \right], \end{aligned} \tag{6}$$

where the first term refers to the deep input gradient. It is noted that optimization with regard to the displacement vector is non-convex due to the high non-linearity of DBP, and thus is conducted in an iterative manner. In particular, we follow the gradient descent method to iteratively update the displacement vector, which is initialized as an all-zero vector. In each inference iteration, the gradient w.r.t. $\Delta\boldsymbol{T}_i$ is computed to transform the present facial landmark structure. Finally, the inferred displacement is acquired as the combination of the input gradients derived in all the iterations.

After we obtain the displacement vectors, standard image warping techniques are applied to produce the beautified image.

3. EXPERIMENT

We compare our face beautification method with the work in [2]. The proposed method is more natural and personalized compared with SVR-beauty. SVR-beauty tends to morph the landmarks towards the mean shape, while our method learns an adaptive transformation for a different shape vector. As the standard to rate the level of attractiveness is vague, 26 volunteers are invited to quantitatively rate the beautified results of our method and SVR-beauty. According to the statistics of the user study, most beautified images show aesthetic improvements over the original copies for our method. In comparison, the number for SVR-beauty is lower. The volunteers also give a higher rating of our method in terms of naturalness: the inferred shape transformations are generally considered as smooth and natural. In summary, the proposed framework gives satisfying performance for facial beautification based on the evaluation from the users.

4. CONCLUSION

In this demo, we showcase an automatic face beautification framework. It performs the beautification task in a learning-based manner and generates personalized and natural beautification results. The beautification of human faces within videos also has very promising application potential. In future, we shall further extend the proposed framework to handle beautification tasks in videos.

5. REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.

[2] T. Leyvand, D. Cohen-Or, G. Dror, and D. Lischinski. Data-driven enhancement of facial attractiveness. *TOG*, 2008.

[3] L. Liu, J. Xing, S. Liu, H. Xu, X. Zhou, and S. Yan. Wow! you are so beautiful today! *ACM TOMM*, 2014.

[4] T. V. Nguyen, S. Liu, B. Ni, J. Tan, Y. Rui, and S. Yan. Sense beauty via face, dressing, and/or voice. In *ACM MM*, 2012.