Transferring Rich Deep Features for Facial Beauty Prediction

Lu Xu, **Jinhai Xiang** and Xiaohui Yuan
Huazhong Agricultural University

CONTENTS

01 Introduction

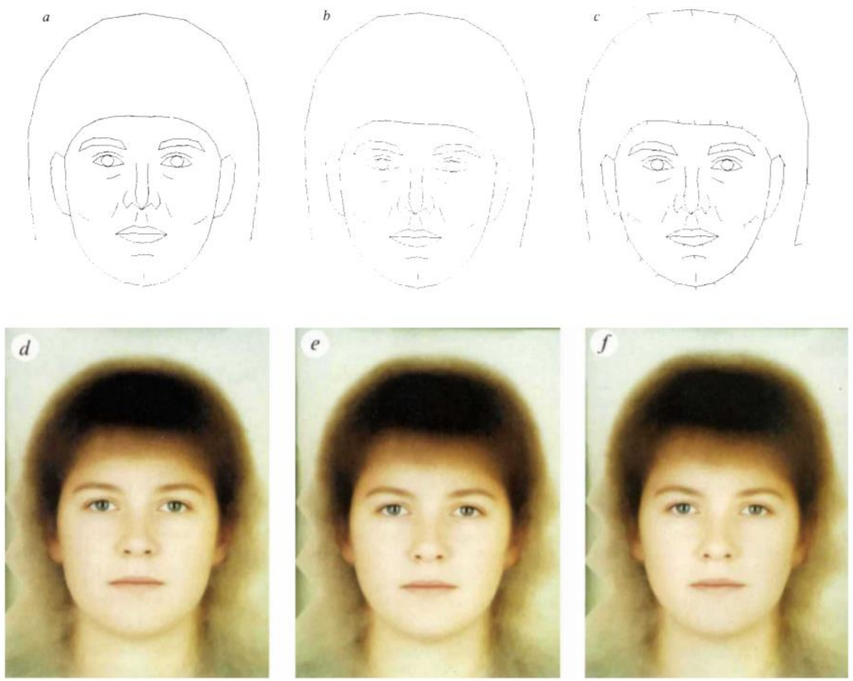
Deep Learning

Deep learning (deep neural network) is a subset of representation learning. It can extract more abstract and more discriminative features than hand-crafted descriptors. DL has been widely used in many fields (especially AI-complete tasks like CV, NLP, and Speech).

Facial Beauty Prediction

FBP aims to recognize facial attractiveness from a portrait image. It has been widely used among SNS and short video platforms (like TikTok, Meitu and Instagram).

Facial Beauty Prediction



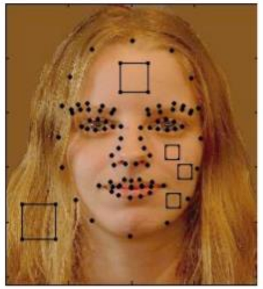


Figure 1: Facial coordinates with hair and skin sample regions as represented by the facial feature extractor. Coordinates are used for calculating geometric features and asymmetry. Sample regions are used for calculating color values and smoothness. The sample image, used for illustration only, is of T.G. and is presented with her full consent.

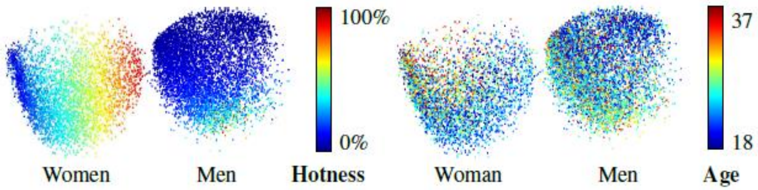


Figure 9. Visualization of latent space Q for women and men.

- **Conclusion**

Facial beauty perception is subjective in personal view, but shows stability in group.

It can be automatically predicted by data-driven models.

- **Tendency**

Moving from hand-crafted features to deep learning models.

1. Perrett, David I., Karen A. May, and Sin Yoshikawa. "Facial shape and judgements of female attractiveness." Nature 368.6468 (1994): 239.
2. Kagian, A., Dror, G., Leyvand, T., Cohen-Or, D., Ruppin, E.: A humanlike predictor of facial attractiveness. NIPS, pp. 649–656 (2007)
3. Rothe, R., Timofte, R., Van Gool, L.: Some like it hot-visual guidance for preference prediction. In: Proceedings CVPR 2016, pp. 1–9 (2016)

02 Related Works

02 Related Works

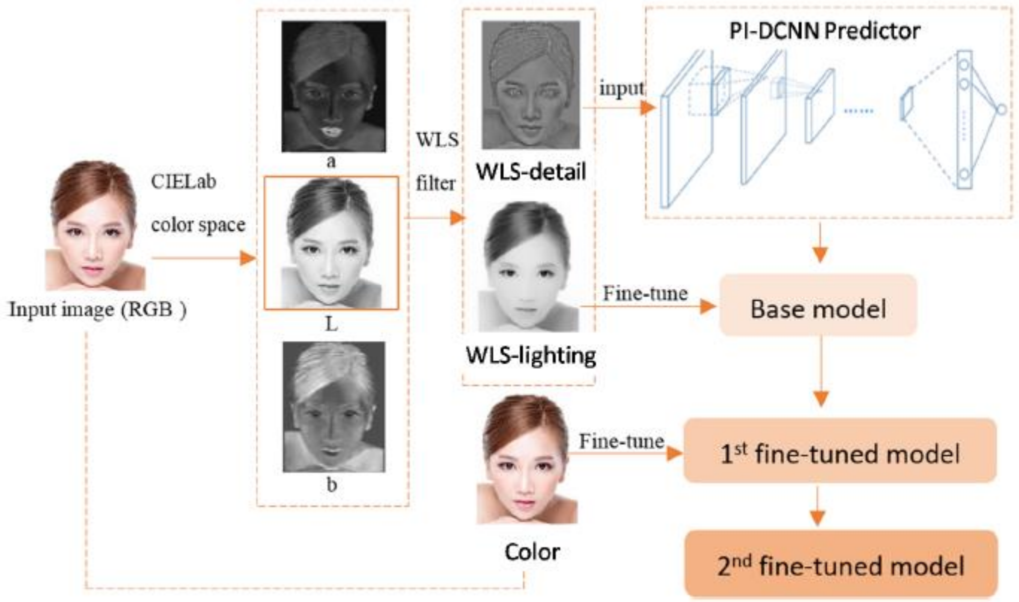


Fig. 3. The cascaded fine-tuning process of PI-CNN using the facial feature extracted by the adaptive WLS filter [21].

In the PI-CNN, RGB channels are used as the color feature, and the detail and lighting features are extracted by an edge-aware filter. Specifically, we used the adaptive WLS filter to decompose the input into separated layers of facial lighting and detail feature, i.e. WLS-lighting and WLS-detail.

Then, we pre-train the PI-CNN predictor using WLS-detail, and sequentially fine-tune the pre-trained model with WLS-lighting and RGB

Xu, Jie, et al. "Facial attractiveness prediction using psychologically inspired convolutional neural network (PI-CNN)." *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*. IEEE, 2017.

02 Related Works

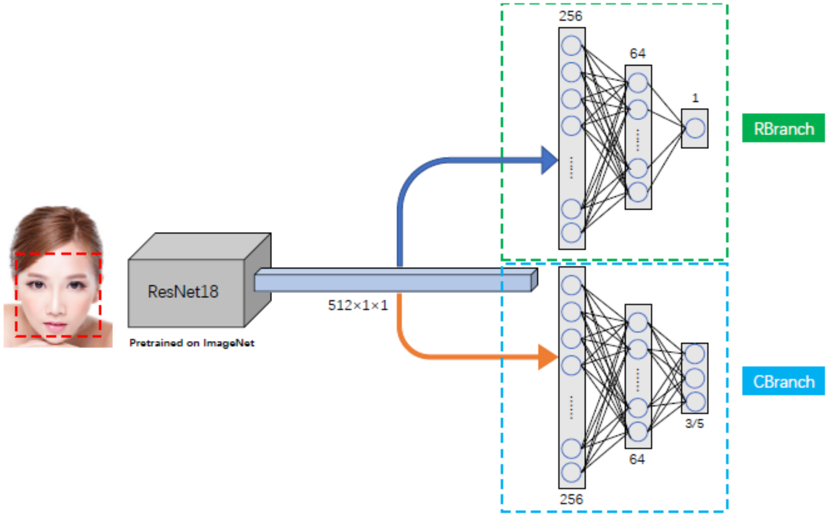


Fig. 1. Architecture of CRNet. ResNet18 [21] is adopted in our feature extraction procedure, the network is split into two branches. RBranch minimizes MSE loss, and CBranch minimize cross entropy loss.

$$\mathcal{L}_c = -\frac{1}{M}\sum_{i=1}^{M} y_i \cdot log\hat{y_i}$$

$$\mathcal{L}_r = \frac{1}{M}\sum_{i=1}^{M}(y_i - \hat{y_i})^2$$

$$\mathcal{L} = \theta_c \cdot \mathcal{L}_c + \theta_r \cdot \mathcal{L}_r$$

Xu, Lu, Jinhai Xiang, and Xiaohui Yuan. "CRNet: Classification and Regression Neural Network for Facial Beauty Prediction." *Pacific Rim Conference on Multimedia (PCM)*. Springer, Cham, 2018.

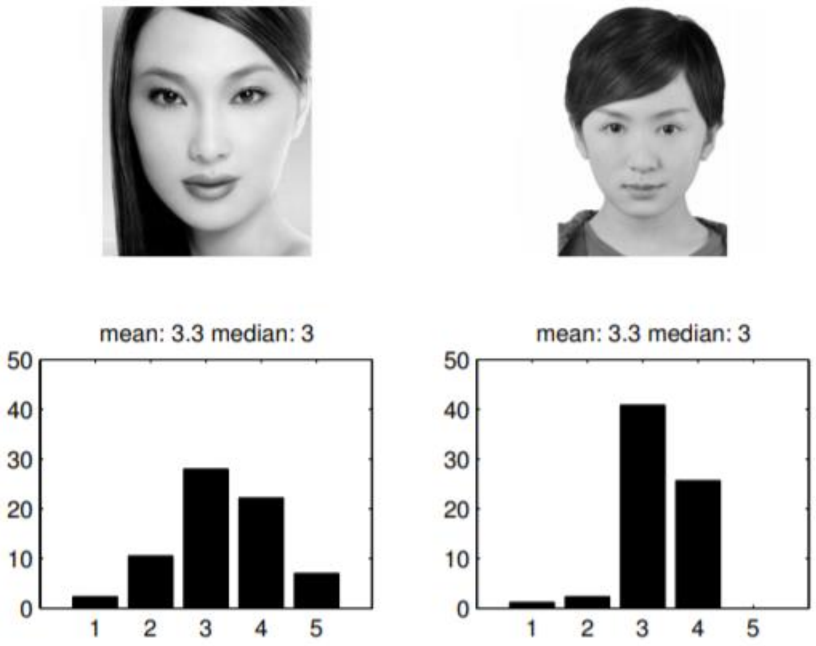


Figure 1: Two images with ratings. The histograms show the number of the raters giving the corresponding ratings.

- Apply label distribution to represent the human sense of beauty, which matches the nature of the subjective human sense better.
- Propose a novel method to covert the k-wise comparisons to label distributions.
- Propose Structural Label Distribution Learning (SLDL) based on structural SVM to predict human sense toward facial beauty.

Ren, Yi, and Xin Geng. "Sense beauty by label distribution learning." *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI)*. 2017.

03 Proposed Method

03 Proposed Method

- Transferring Rich Deep Features for Facial Beauty Prediction (TransFBP)

Feature matters in approximately every computer vision task. But deep learning models are easy to stuck in overfitting, and often need lots of labeled data to train.

So can we leverage existing dataset from <u>different tasks</u> to form facial beauty representation?

- Transferring Rich Deep Features for Facial Beauty Prediction (TransFBP)

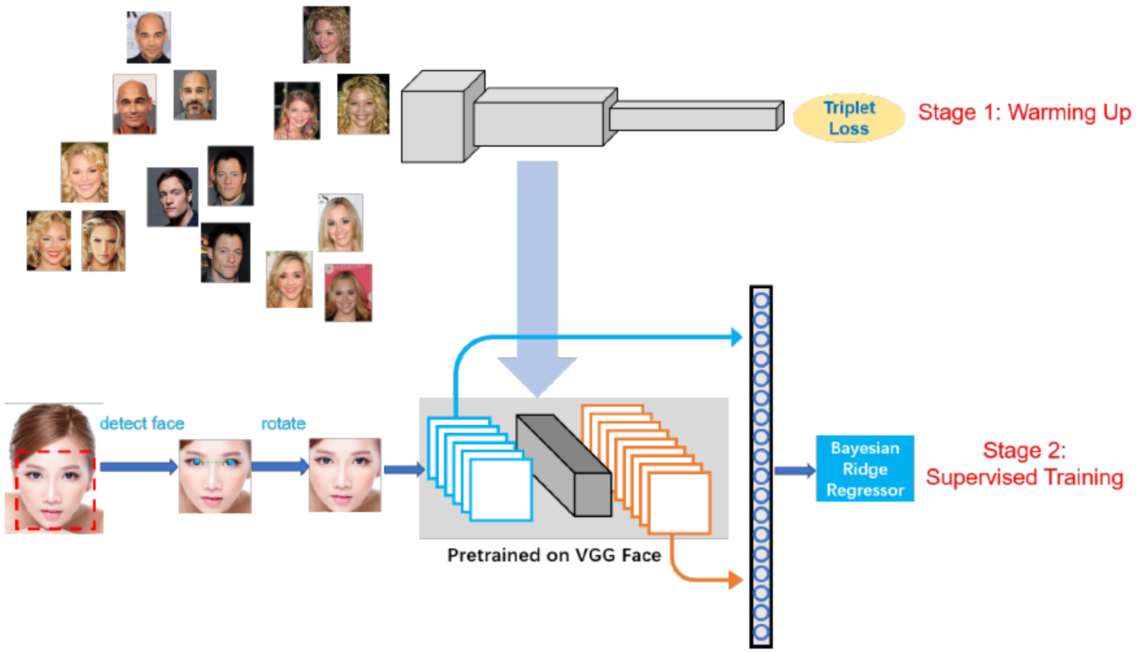


Fig. 1. Pipeline of our proposed method. We firstly train a face verification task on VGG Face dataset to obtain facial beauty representation. Then the face is detected, rotated and then fed into the pre-trained model, we concatenate both low level and high-level features for more informative facial representation, and flatten them into feature vectors for the input of Bayesian ridge regression.

$$E(W') = \sum_{(a,p,n)\in T} max\{0, \alpha - ||x_a - x_n||_2^2 + ||x_a - x_p||_2^2\}, \quad x_i = W' \frac{\phi(l_i)}{||\phi(l_i)||_2}$$

A two-stage model was proposed in our paper.

Namely, we first conduct large-scale pretraining with *Triplet Loss* on a *Face Verification task* for *Warming Up*.
Then the facial representations from both lower and higher layers are concatenated to fit *Bayesian Ridge Regression*.

- Transferring Rich Deep Features for Facial Beauty Prediction (TransFBP)

The estimation of parameter β in ridge regression [22] can be obtained by L_2-regularized least squares,

$$\hat{\beta} = \underset{\beta}{argmin}(||y - \mu\mathbf{1} - X\beta||^2 + \lambda||\beta||^2) \tag{2}$$

where $y = \{y_1, \cdots, y_n\}$, $\mu\mathbf{1}$, X and $\beta = \{\beta_1, \cdots, \beta_n\}$ represent the label, bias, feature and learned weights, respectively.

A Bayesian view of ridge regression can be obtained by minimizing Equation (2) as the posterior mean of a model where $\beta_j \sim \mathcal{N}(0, \sigma^2/\lambda)$ for all j, where σ^2 represents variance. In our experiments, we use an unscaled Gaussian prior [27] for the coefficients: $\beta_j \sim \mathcal{N}(0, 1/\lambda)$ for all j.

The output y is assumed to be Gaussian distributed around $X\beta$ in order to form a fully probabilistic model:

$$P(y|X, \beta, \alpha) = \mathcal{N}(y|X\beta, \alpha) \tag{3}$$

Bayesian ridge regressor evaluates a probabilistic model of the regression problem. The prior for the parameter β is decided by a spherical Gaussian:

$$P(\beta|\lambda) = \mathcal{N}(\beta|0, \lambda^{-1}I_p) \tag{4}$$

The priors over α and λ are chosen to be Gamma distributions, the conjugate prior for the precision of the Gaussian. The parameters β, α and λ are estimated jointly during the fit procedure. All the parameters are learned by maximizing the marginal log-likelihood. The pipeline of our method is shown in Fig. 1.

03 Proposed Method

- Transferring Rich Deep Features for Facial Beauty Prediction (TransFBP)

Fig. 3. Precisely predicted samples (up), and imprecisely predicted samples (bottom) in ECCV HotOrNot dataset. The images with diverse facial expression and variant postures are hard to predict by our method.

- Transferring Rich Deep Features for Facial Beauty Prediction (TransFBP)

Table 3. Performance comparison with other methods. Our method achieves state-of-the-art performance on the SCUT-FBP dataset. The best performance is highlighted in bold.

Method	PC
Combined Features+Gaussian Reg [3]	0.6482
CNN-based [3]	0.8187
Liu et al. [23]	0.6938
KFME [25]	0.7988
RegionScatNet [26]	0.83
PI-CNN [11]	0.87
Ours	**0.8742**

Table 5. Performance comparison with other state-of-the-art on ECCV HotOrNot dataset [4]. Our method achieves the best result.

Method	PC
Eigenface [4]	0.180
Multiscale Model [4]	0.458
Auto Encoder [10]	0.437
Ours	**0.468**

03 Proposed Method

- Transferring Rich Deep Features for Facial Beauty Prediction (TransFBP)

◆ It indicates that the model pretrained on a totally <u>different task</u> with <u>different data distribution</u>, guided by <u>different loss function</u> also contains informative representation for beauty.

◆ Our proposed two-stage models achieves state-of-the-art performance on relevant benchmark datasets.

◆ The features can be shared among different tasks as well.

03 Proposed Method

- Transferring Rich Deep Features for Facial Beauty Prediction (TransFBP)

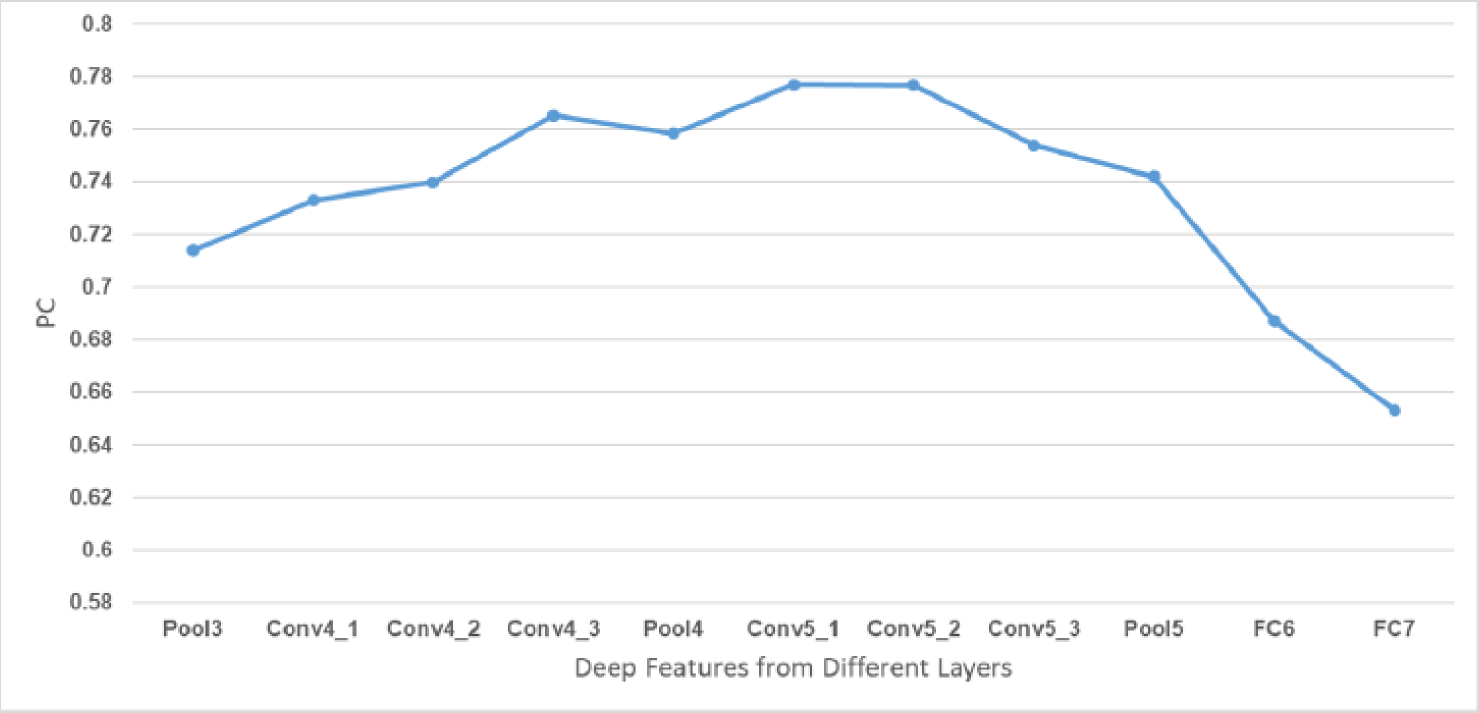


Figure 4: Facial beauty representation power in “warming up strategy” of different layers.

As layer goes deeper, we can see a clear performance drop. It can be explained that higher layers contain more task-specific information.

04 Future Works

- Since features can be shared among different tasks, we are developing multi-task models for face-related tasks.
- Adopt more powerful loss functions as supervision.

Transferring Rich Deep Features for Facial Beauty Prediction

Lu Xu, **Jinhai Xiang** and Xiaohui Yuan
Huazhong Agricultural University