Hierarchical Multi-Task Network For Race, Gender and Facial Attractiveness Recognition

Speaker: Jinhai Xiang
Huazhong Agricultural University

CONTENTS

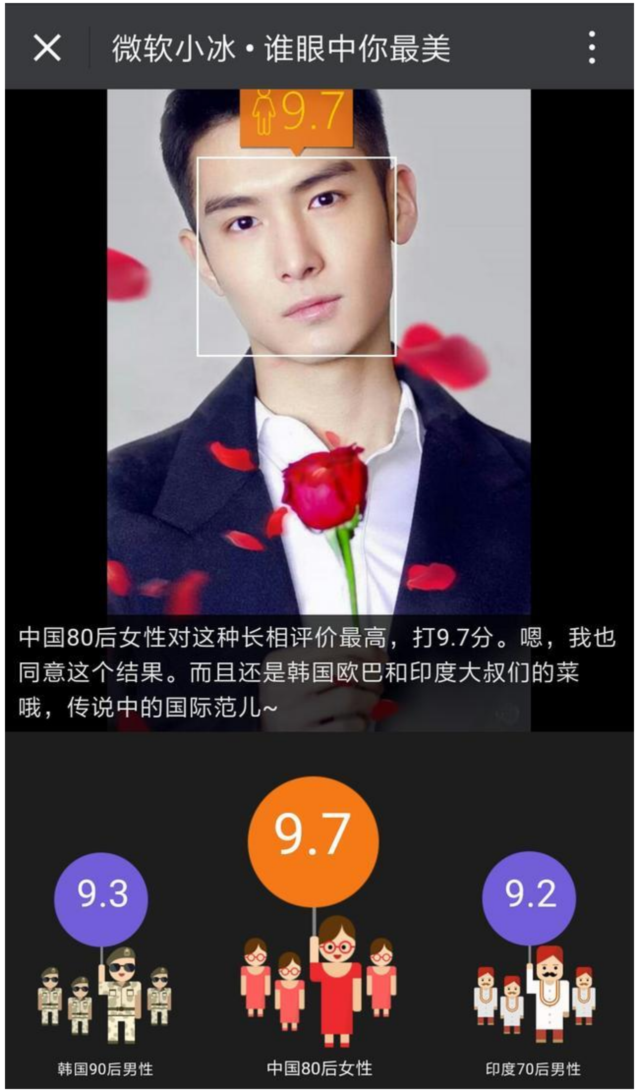
01 Introduction

Facial Beauty Prediction

Pursing beauty is the nature of human beings.

Ancient times

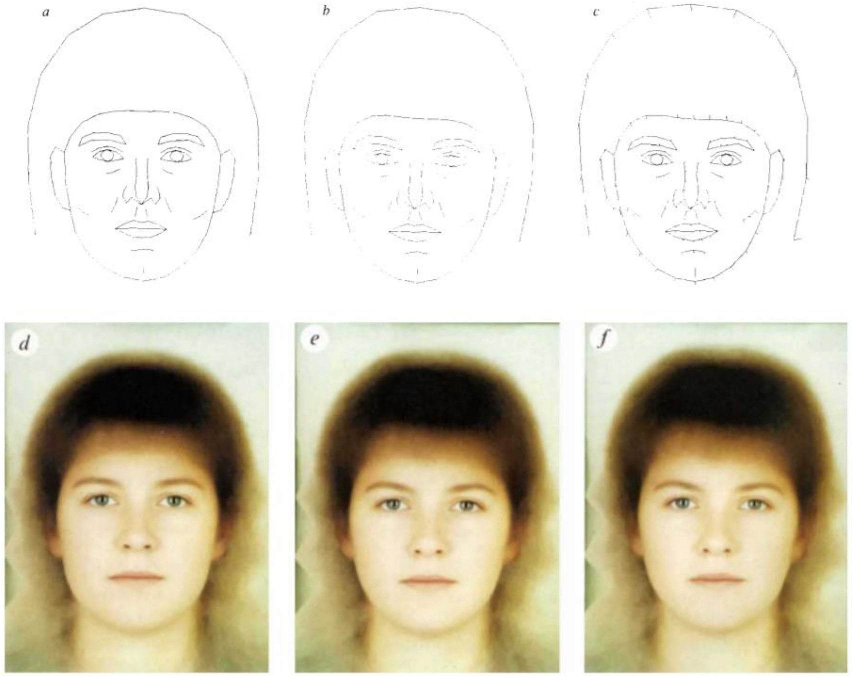
Nowadays



FBP in TikTok

FBP in Microsoft XiaoBing

Facial Beauty Prediction



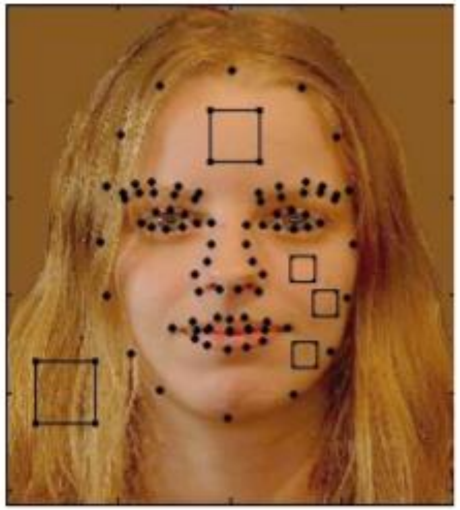


Figure 1: Facial coordinates with hair and skin sample regions as represented by the facial feature extractor. Coordinates are used for calculating geometric features and asymmetry. Sample regions are used for calculating color values and smoothness. The sample image, used for illustration only, is of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

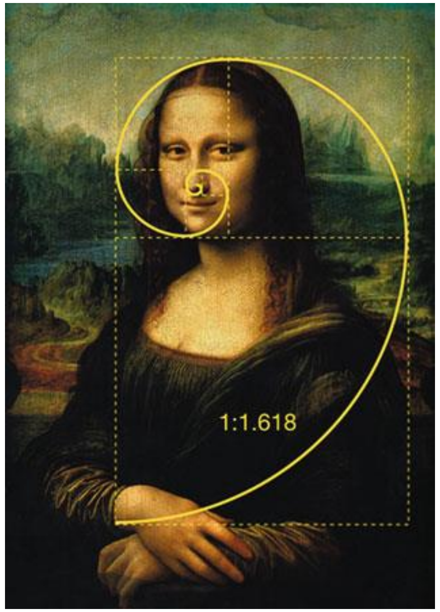
Facial attractiveness can be automatically predicted by data-driven models.

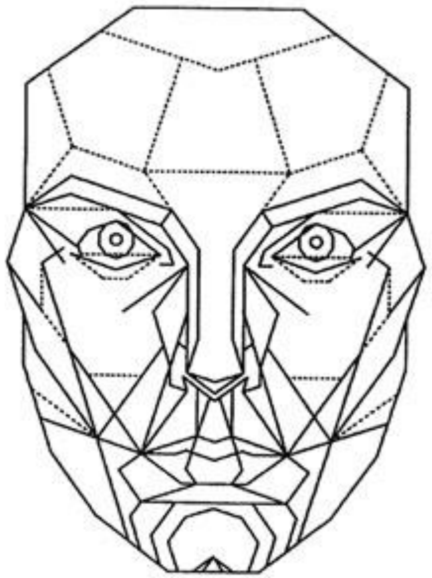
1. Perrett, David I., Karen A. May, and Sin Yoshikawa. "Facial shape and judgements of female attractiveness." **Nature** 368.6468 (1994): 239
2. Kagian, A., Dror, G., Leyvand, T., Cohen-Or, D., Ruppin, E.: A humanlike predictor of facial attractiveness. **NIPS**, pp. 649–656 (2007)
3. Rothe, R., Timofte,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

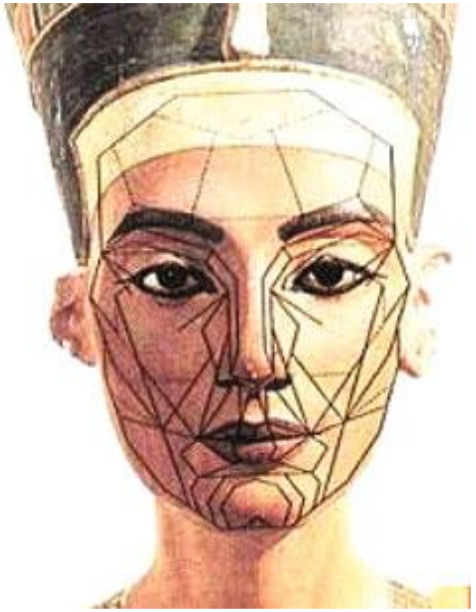
Facial Beauty Prediction

Typical Facial Beauty Analysis

- **Golden Ratio Rules**





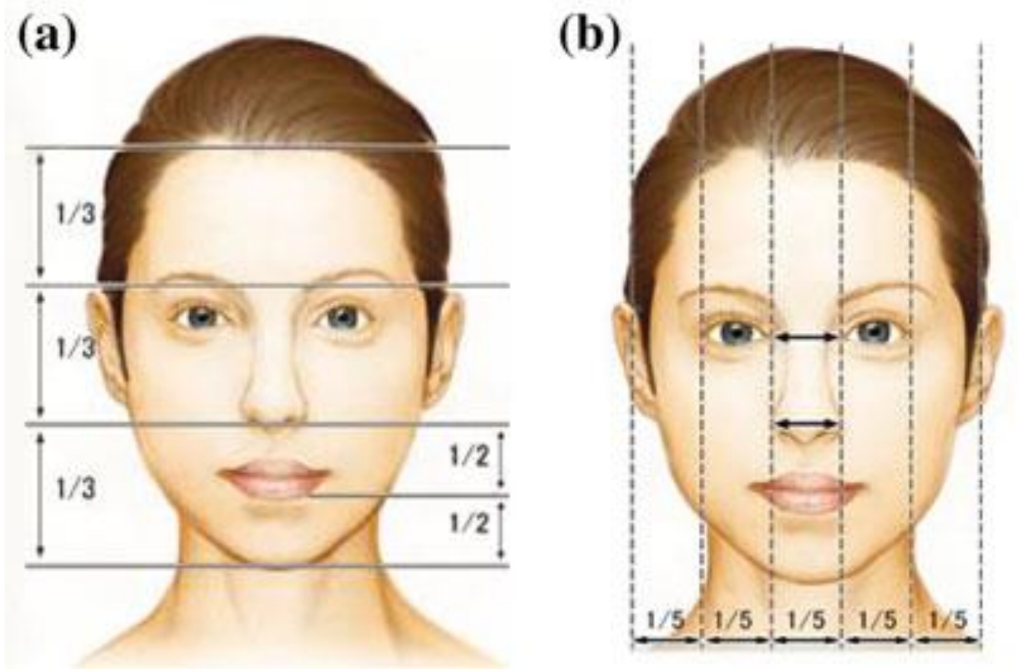


1. David Zhang, Fangmei Chen, Yong Xu et al., Computer models for facial beauty analysis, Springer, 2016.

Facial Beauty Prediction

Typical Facial Beauty Analysis

- **Vertical Thirds and Horizontal Fifths**



1. David Zhang, Fangmei Chen, Yong Xu et al., Computer models for facial beauty analysis, Springer, 2016.

Facial Beauty Prediction

Typical Facial Beauty Analysis

- **Averageness Hypothesis**

Average face is more attractive than component face.

1. David Zhang, Fangmei Chen, Yong Xu et al., Computer models for facial beauty analysis, Springer, 2016.

Facial Beauty Prediction

Typical Facial Beauty Analysis

Hand-crafted Features + Machine Learning Classifier/Regressor

- Feature extraction
  - geometry features
  - texture features
  - color features

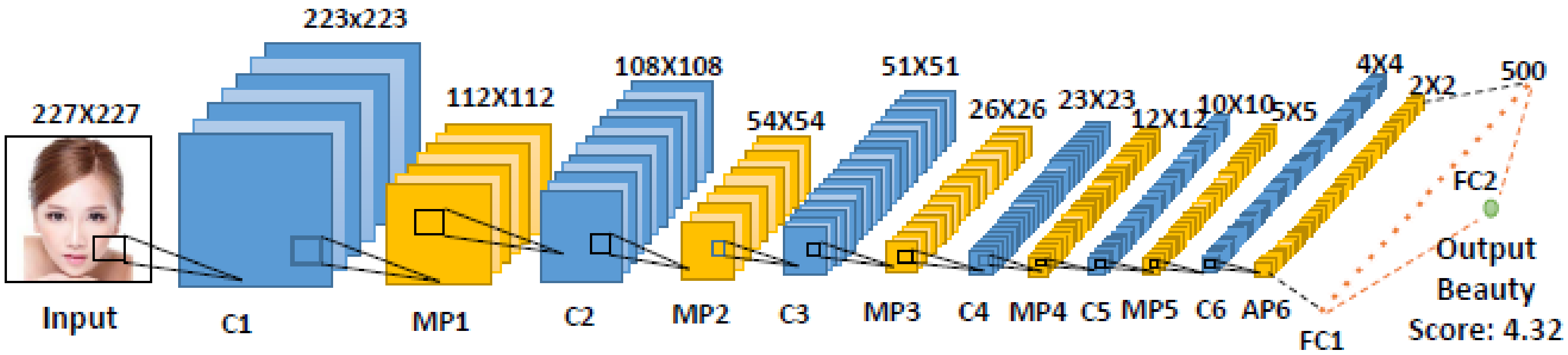
- Modeling methods
  - SVR
  - KNN
  - NN

1. David Zhang, Fangmei Chen, Yong Xu et al., Computer models for facial beauty analysis, Springer, 2016.

Facial Beauty Prediction

Recent Advance in Facial Beauty Analysis

- **Deep Learning for facial attractiveness regression**



The architecture of PI-DCNN for facial attractiveness prediction

Moving from hand-crafted features to deep learning.

1. Jie Xu, Lianwen Jin, Lingyu Liang, Ziyong Feng, Duorui Xie, Huiyun Mao, "Facial attractiveness prediction using psychologically inspired convolutional neural network (pi-cnn)", **ICASSP**, 2017.
2. Duorui Xie, Lingyu Liang, Lianwen Jin, Jie Xu, Mengru Li, "Scut-fbp: A benchmark dataset for facial beauty perception", **SMC**, 2015.

Facial Beauty Prediction

Recent Advance in Facial Beauty Analysis

- **Label Distribution Learning**

Algorithm 1 Structural Label Distribution Learning(SLDL)

Input: $D = \{(\boldsymbol{x}_1, \boldsymbol{d}_1), \cdots, (\boldsymbol{x}_N, \boldsymbol{d}_N)\}, C, \epsilon$

$Q \longleftarrow \varnothing$

repeat

  compute $(\boldsymbol{w}, \xi)$ in Eq.(12)

  for $i = 1, \cdots, N$ **do**

    $\hat{\boldsymbol{d}}_i \longleftarrow \arg\max_{\hat{\boldsymbol{d}}_i \in V} \triangle(\boldsymbol{d}_i, \hat{\boldsymbol{d}}_i) + \langle \boldsymbol{w}, \psi(\boldsymbol{x}_i, \hat{\boldsymbol{d}}) \rangle$

  end for

  if $\frac{1}{N}\sum_{i=1}^{N} \triangle(\boldsymbol{d}_i, \hat{\boldsymbol{d}}_i) - \frac{1}{N}\sum_{i=1}^{N} \langle \boldsymbol{w}, \delta\psi_i(\hat{\boldsymbol{d}}_i) \rangle > \xi + \epsilon$ **then**

    $Q \longleftarrow Q \bigcup (\hat{\boldsymbol{d}}_1, \cdots, \hat{\boldsymbol{d}}_N)$

  end if

until Q has no change

return $(\boldsymbol{w}, \xi)$

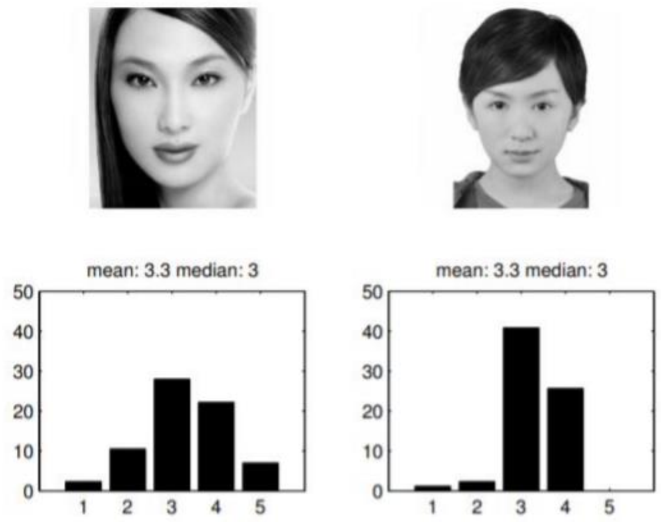


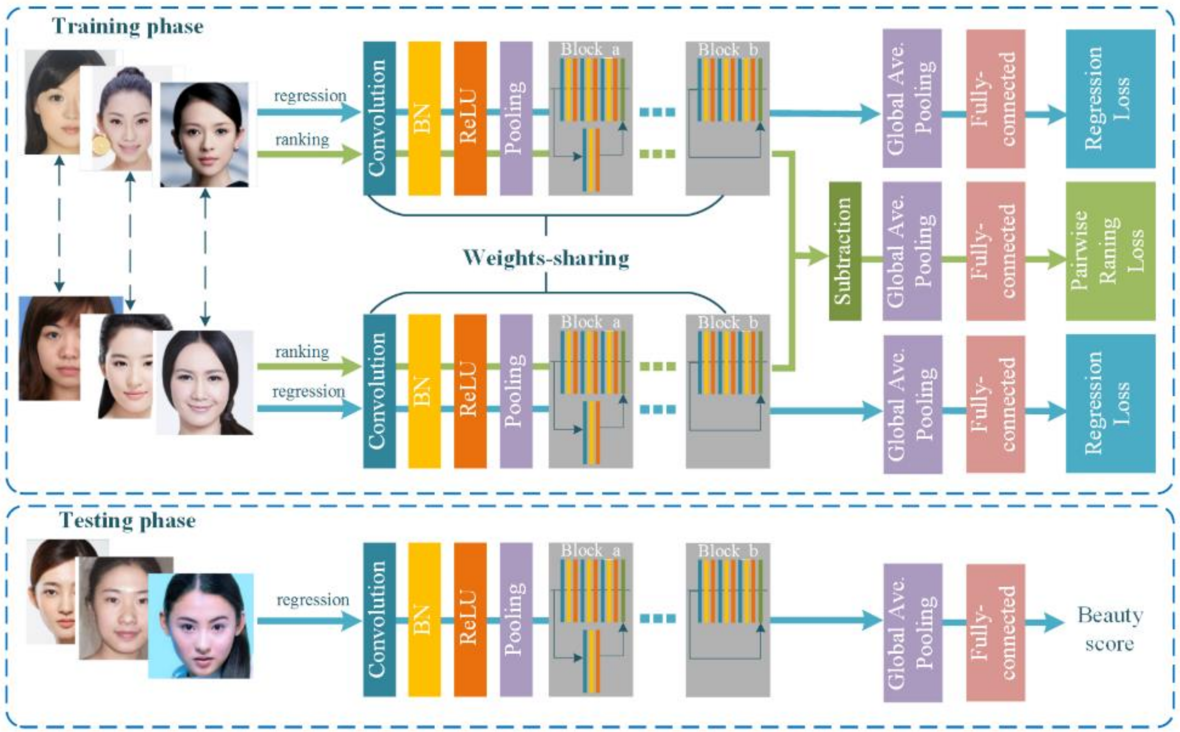
Figure 1: Two images with ratings. The histograms show the number of the raters giving the corresponding ratings.

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

Facial Beauty Prediction

Recent Advance in Facial Beauty Analysis

- **Ranking**

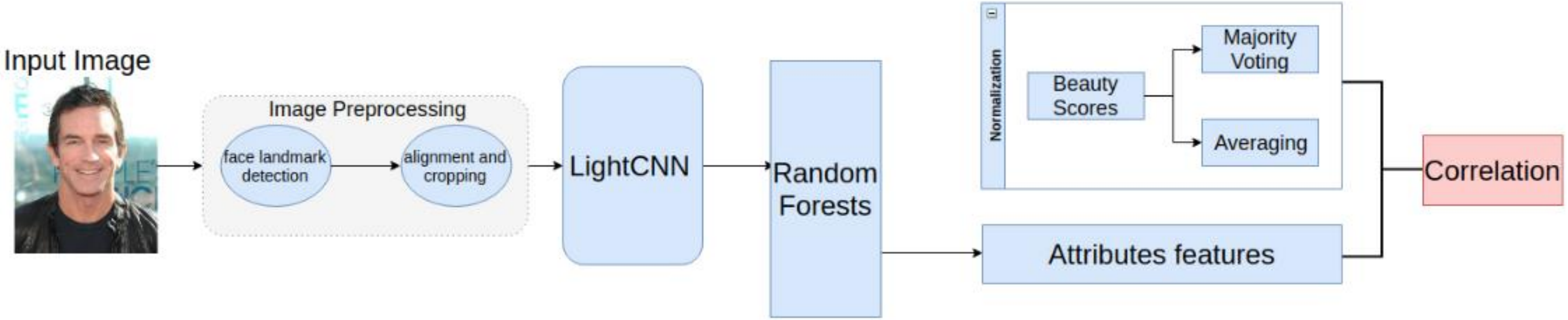


1. Lin, Luojun, Lingyu Liang, and Lianwen Jin. "R 2-ResNeXt: A ResNeXt-Based Regression Model with Relative Ranking for Facial Beauty Prediction." 2018 24th International Conference on Pattern Recognition (**ICPR**). IEEE, 2018.

Facial Beauty Prediction

Recent Advance in Facial Beauty Analysis

- **Analysis on Deep Facial Beauty Prediction Models**



1. Liu, Xudong, et al. "Understanding beauty via deep facial features." Proceedings of the IEEE **Conference on Computer Vision and Pattern Recognition Workshops**. 2019.

Facial Beauty Prediction

Challenges in Facial Beauty Analysis

- Diverse pose
- Expression
- Low resolution
- Different races & genders

It is still quite difficult to develop accurate facial beauty predictors.

1. Douglas Gray, Kai Yu, Wei Xu, Yihong Gong, "Predicting facial beauty without landmarks", **ECCV**, 2010.
2. Lingyu Liang, Luojun Lin, Lianwen Jin, Duorui Xie, Mengru Li, "Scut-fbp5500: A diverse benchmark dataset for multi-paradigm facial beauty prediction", **ICPR**, 2018.

02 Proposed Methods

Components of HMTNet

- How to form discriminative representation?
  Feature Aggregator

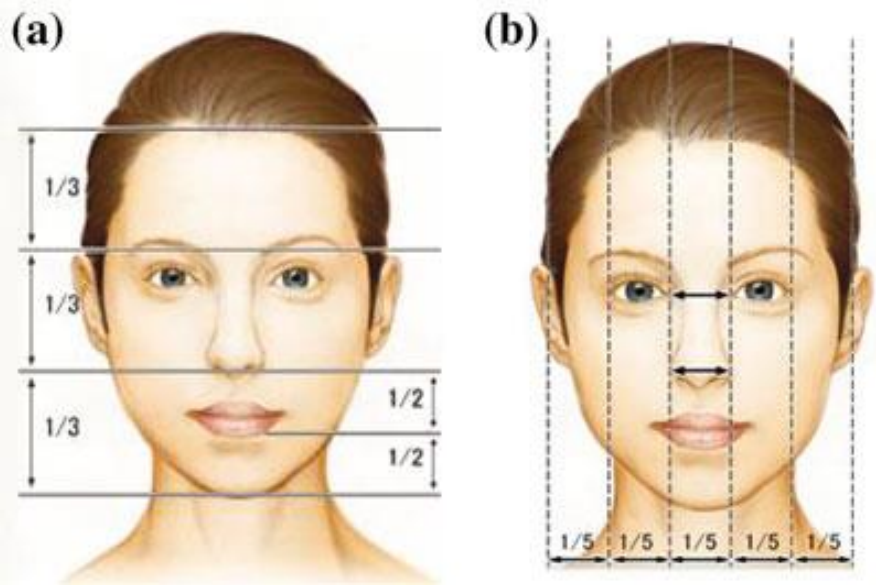
- How to construct task-specific layers?
  Hierarchical Branch Strategy

- How to better supervise model training?
  Smooth Huber Loss

Feature Aggregator

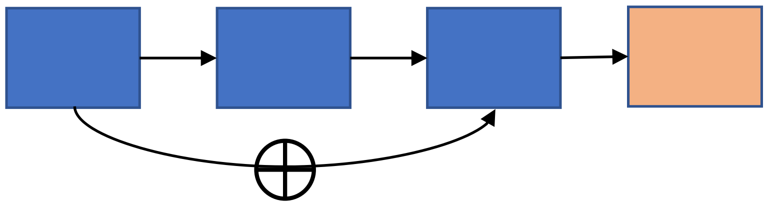
- Low-level features contain more detailed information, such as blob, texture, etc.
- High-level representation embeds rich semantic meaning.

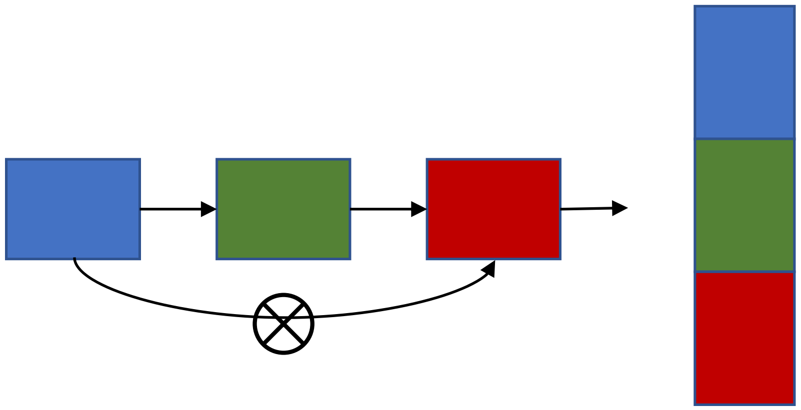
- Low-level information (such as facial geometric information) also contributes to beauty perception.



Recall from Vertical Thirds and Horizontal Fifths (“三庭五眼” in Chinese).

Feature Aggregator





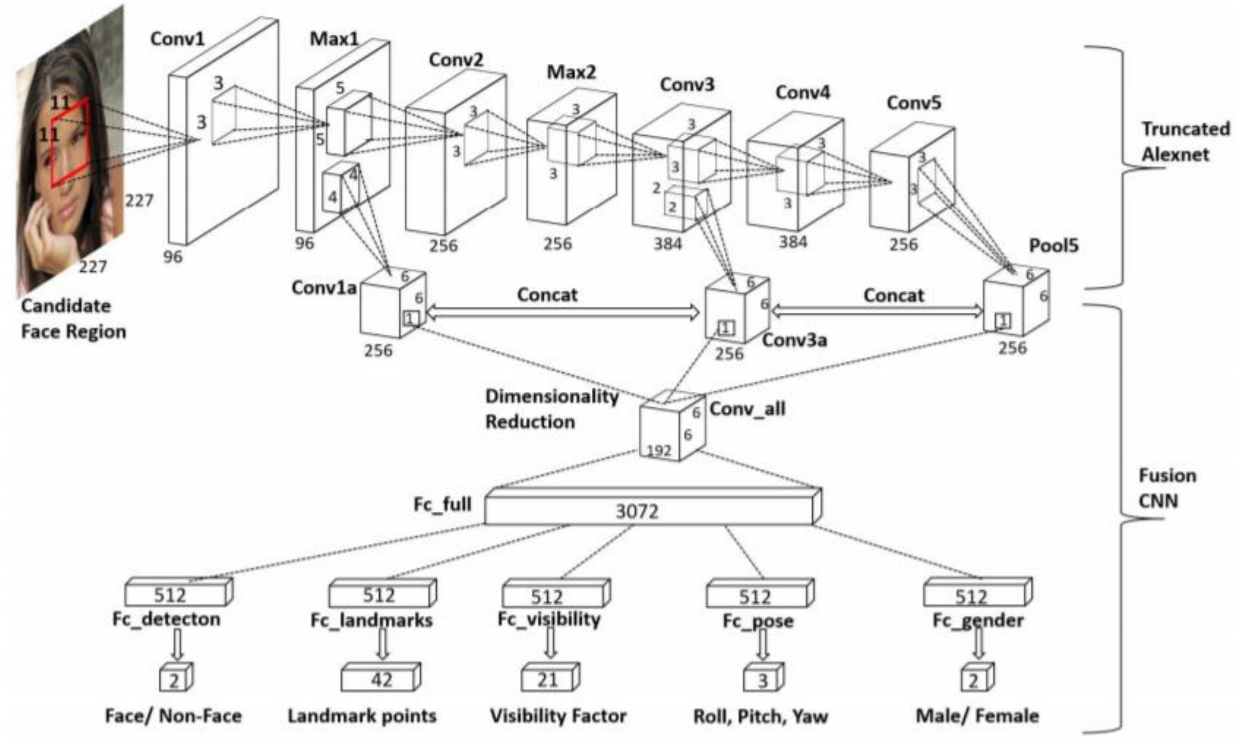
$$f_{avg} = \frac{1}{\mathcal{C}} \sum_{i=1}^{\mathcal{C}} fm_i, \quad fm_i, f_{avg} \in \mathbb{R}^{w \times h \times c} \qquad (1)$$

$$f_{concat} = fm_1 \otimes \cdots \otimes fm_{\mathcal{C}}, \quad f_{concat} \in \mathbb{R}^{w \times h \times c \times \mathcal{C}} \qquad (2)$$

Feature aggregator can form more discriminative representation by aggregating features from different layers for diverse recognition tasks.

Hierarchical Branch Strategy

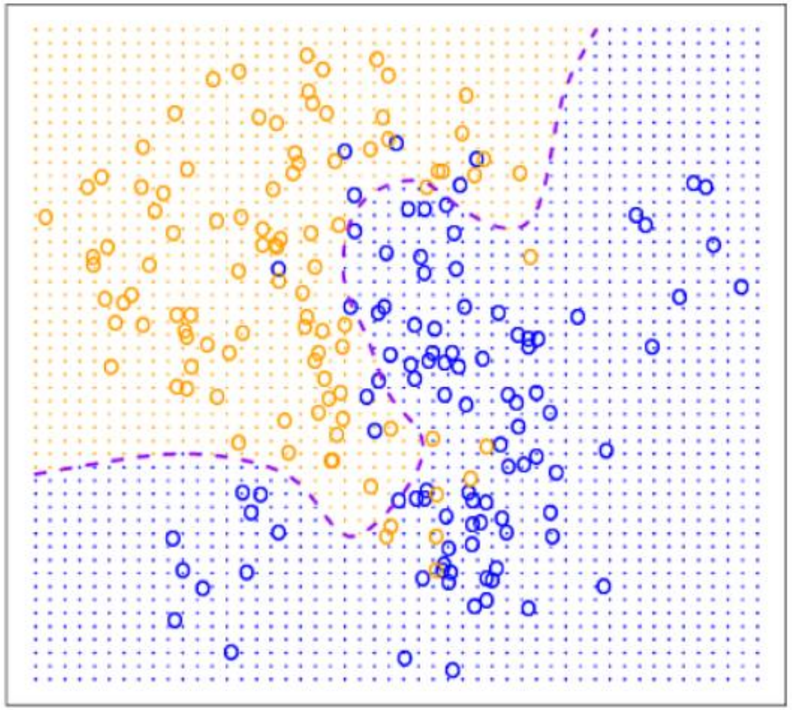
Existing multi-task models often reuse the features in the last layers directly [22, 24, 23].

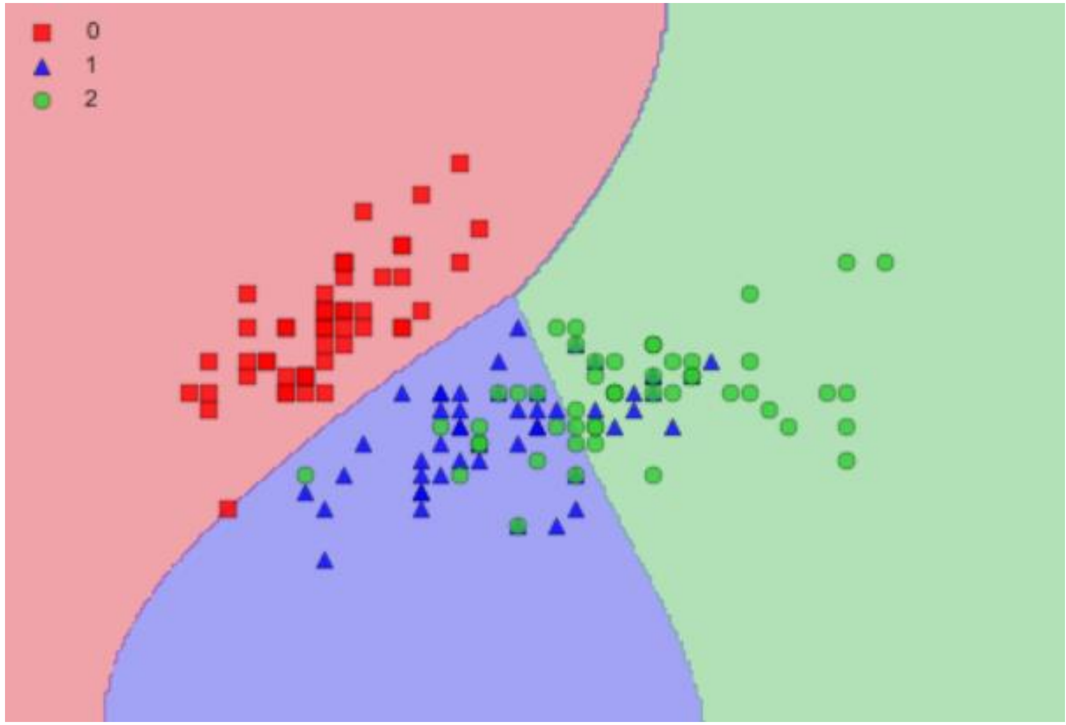


1. Ranjan, Rajeev, Vishal M. Patel, and Rama Chellappa. "Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition." IEEE **Transactions on Pattern Analysis and Machine Intelligence** 41.1 (2017): 121-135.

Hierarchical Branch Strategy

Branch out task-specific layers by
the learning difficulty of decision boundary.





Hierarchical Branch Strategy

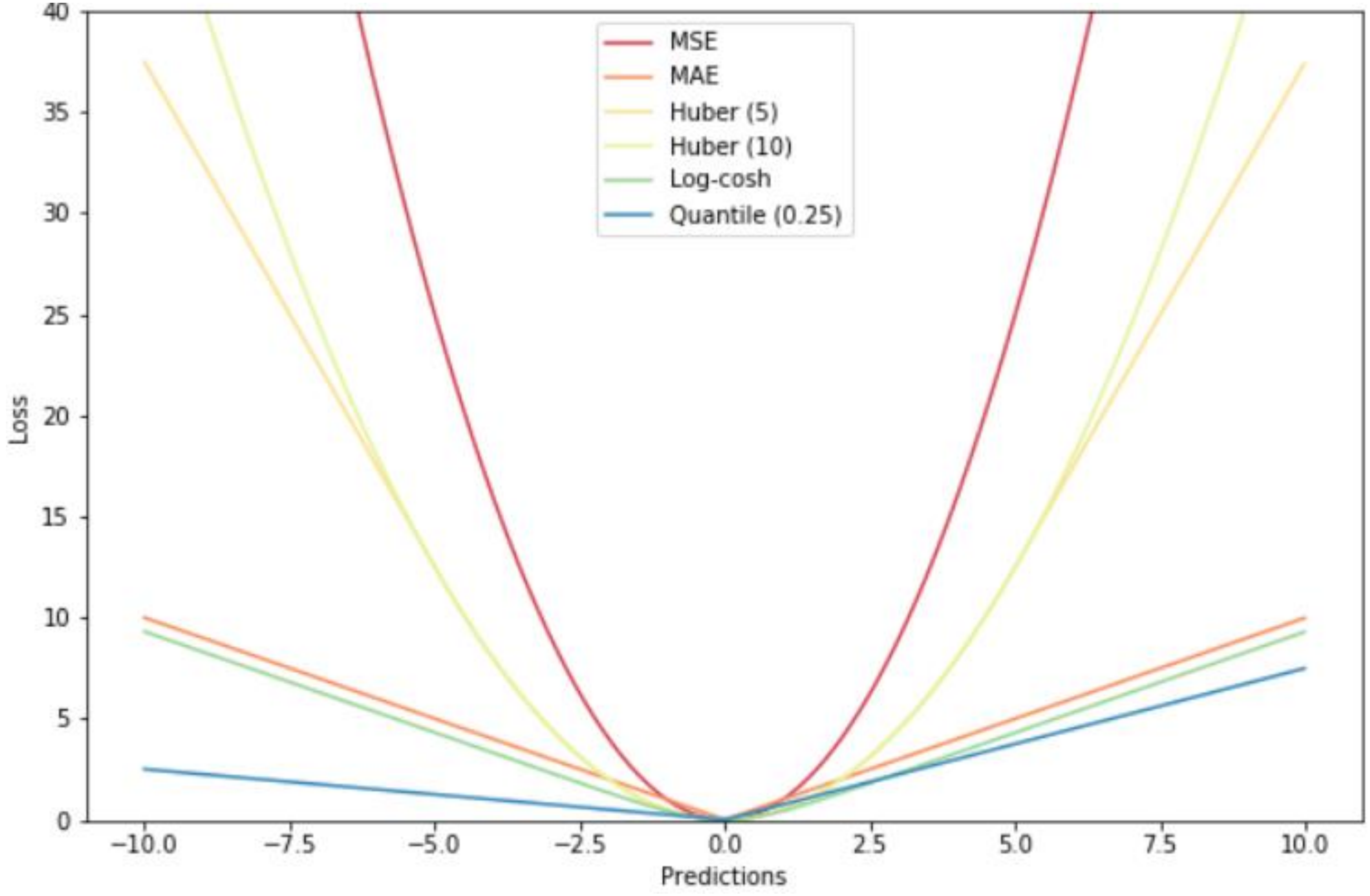
We take a different fashion for branch strategy. Namely, the sub-networks for relative easier tasks are branched out in relative lower layers and embed coarser information, while the sub-networks for difficult tasks are branched out in relative higher layers.

The advantages of this strategy is that we can not only form more informative and richer representations, but also reduce the computational burden as well.

Optimization Object

- MSE Loss is widely used in conventional facial beauty prediction task.
- MSE Loss gives a more stable and closed form solution, but it is easily influenced by outliers.
- MAE Loss is more robust to outliers, but its derivatives are not continuous.

Optimization Object



https://heartbeat.fritz.ai/5-regression-loss-functions-all-machine-learners-should-know-4fb140e9d4b0

Optimization Object

$$Loss_g = -g log(\hat{g}) - (1-g) log(1-\hat{g})$$

$$Loss_r = -\sum_i r_i log(\hat{r}_i)$$

$$\boxed{Loss_a = \begin{cases} \sum_i log(cosh(a_i - \hat{a}_i)) & if |a_i - \hat{a}_i| \le \delta \\ \sum_i \delta |a_i - \hat{a}_i| - \frac{1}{2}\delta^2 & otherwise \end{cases}}$$

$$Loss_{all} = \sum_{t \in \{g,r,a\}} \alpha_t Loss_t$$

- We introduce a new loss function for solving FBP task, which is called "Smooth Huber Loss". It follows a Huber fashion, but it's smoother, and is more robust to outliers.
- It achieves best performance compared with MSE, MAE and Smooth L1 Loss. (see Ablation Analysis).

Architecture of HMTNet

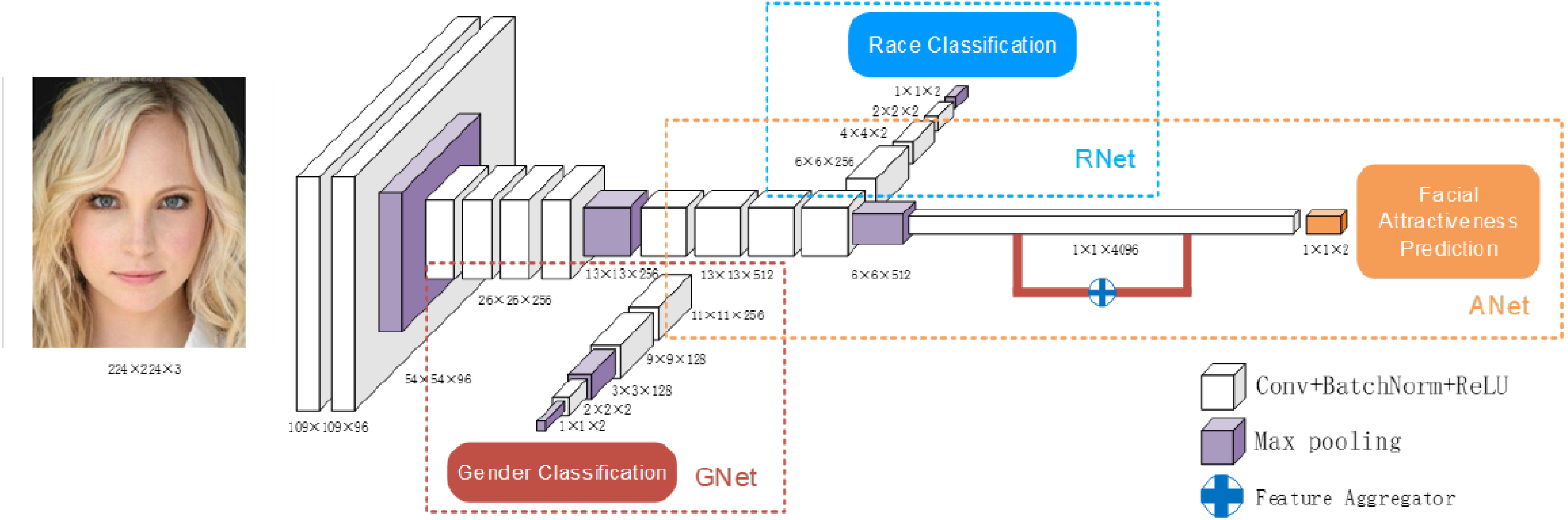


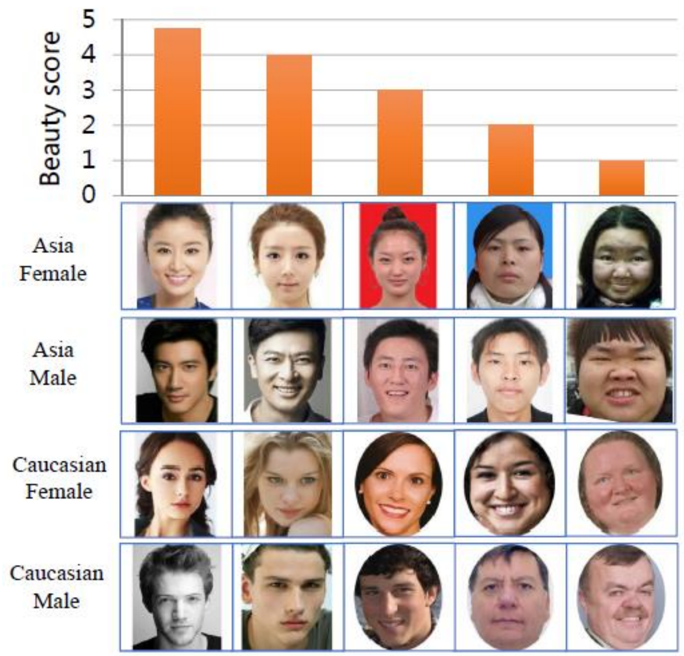
Fig. 1: Overall architecture of HMT-Net. The portrait image is resized to 224 × 224 pixels with RGB channels, and then fed into HMT-Net. RNet and GNet are used to recognize the race and gender, respectively. ANet is utilized to predict the facial attractiveness score. The parameters in low layers can be shared among three sub-networks (GNet, RNet and ANet). All the layers are fully convolutional, and all three branched layers are trained jointly.

03 Experiments

Datasets

- SCUT-FBP5500
  5500 portrait images with rich annotation (beauty score, gender, race, etc.)

- SCUT-FBP
  500 Chinese female facial images



Experimental Results

Table 2. Performance comparison on SCUT-FBP5500.

Model	**MAE**	**RMSE**	**PC**
AlexNet [18, 16]	0.2938	0.3819	0.8298
ResNet-18 [19, 16]	0.2818	0.3703	0.8513
ResNeXt-50 [20, 16]	0.2518	0.3325	0.8777
CRNet [26]	0.2835	0.3677	0.8558
HMTNet (Ours)	**0.2501**	**0.3263**	**0.8783**

Ablation Analysis

- **Effects of Multi-task Joint Training**

Table 3. Evaluation on joint training.

With Joint Training			Without Joint Training		
Acc_r	Acc_g	PC	Acc_r	Acc_g	PC
99.26%	**98.16%**	**0.8783**	98.62%	97.56%	0.8616

- **Effects of Smooth Huber Loss**

Table 4. Evaluation on different loss functions.

Loss Function	**MAE**	**RMSE**	**PC**
MSE Loss	0.2556	0.3372	0.8693
L_1 Loss	**0.2500**	0.3299	0.8753
Smooth L_1 Loss	0.2531	0.3313	0.8738
Smooth Huber Loss	0.2501	**0.3263**	**0.8783**

Ablation Analysis

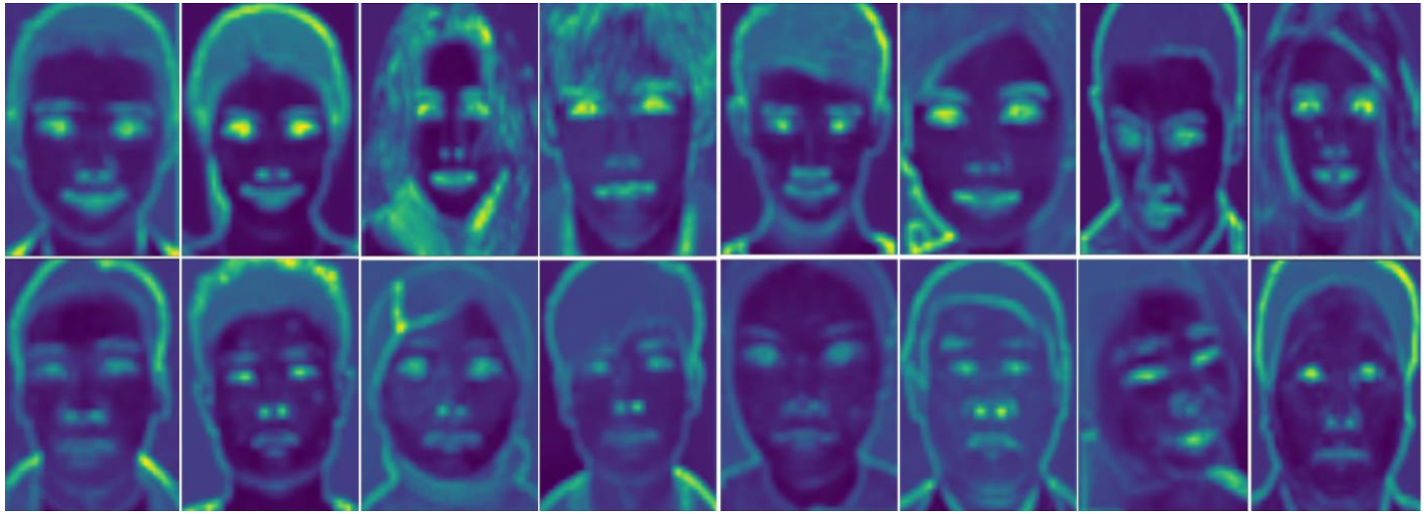
- **Effects of Feature Transferability on Multi-task Training**

Table 5. Performance comparison on SCUT-FBP.

Methods	**PC**
Combined Features+Gaussian Reg [17]	0.6482
CNN-based [17]	0.8187
Liu et al. [27]	0.6938
KFME [28]	0.7988
RegionScatNet [5]	0.83
PI-CNN [6]	0.87
CRNet [26]	0.8723
Ours	**0.8977**

By fitting a simple linear regression model with deep features learned by HMTNet, our method achieves the best performance on SCUT-FBP dataset with a large margin compared with others.

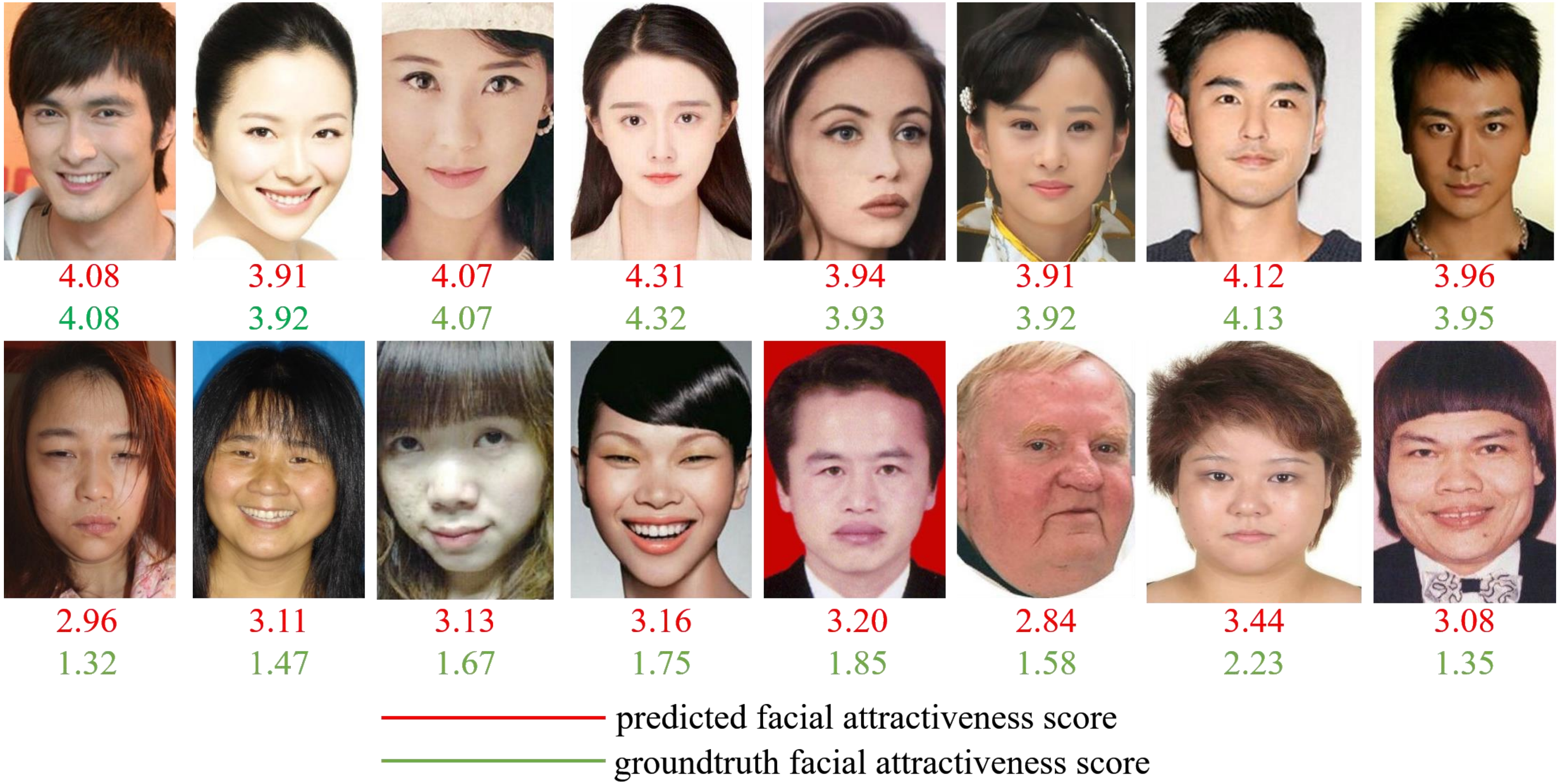
Deep Feature Visualization



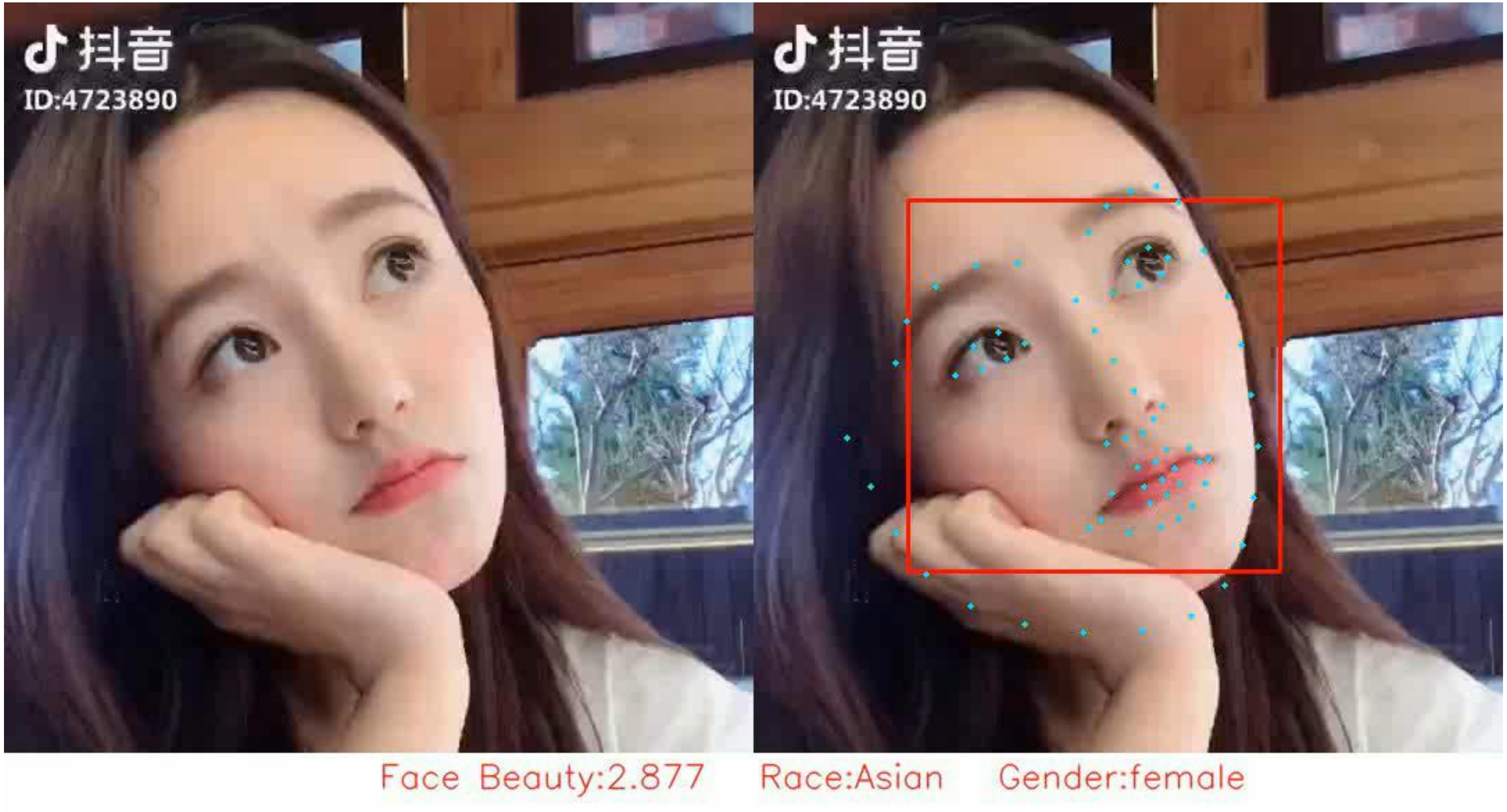
We list both precisely predicted and imprecisely predicted images. Surprisingly and interestingly, HMTNet seems to show more bias on attractive faces since the predicted values of attractive faces are more accurate than those with unattractive faces.

We can see that eyes play a significant role in facial beauty perception. The fashionable hairstyle also contribute to beauty impression.

02 Proposed Methods



Analyze Your Face in Real-time



04 Conclusion and Future Works

Conclusion

- We propose a novel multi-task network with fully convolutional architecture named HMTNet, to simultaneously recognize a person's gender, race and facial beauty score with very promising results.
- We introduce a useful loss function in FBP task for beauty attractiveness regression.

Future Works

- Integrate more sub-tasks (such as facial landmark localization) to embed discriminative geometry information in MTL model.
- Explore more advanced regression loss functions to supervise model training.
- Adopt GANs to generate more portrait images to enhance model learning.
- Explore NAS to search more advanced architecture.

Q&A

Speaker: Jinhai Xiang
Huazhong Agricultural University

Thanks

Speaker: Jinhai Xiang
Huazhong Agricultural University