The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

# FR-ANet: A Face Recognition Guided Facial Attribute Classification Network

**Jiajiong Cao,[†,1] Yingming Li,[*1] Xi Li,[2] Zhongfei Zhang[1]**
[1] College of Information Science & Electronic Engineering, Zhejiang University
[2] College of Computer Science and Technology, Zhejiang University
No.38 Zheda Road, Hangzhou, Zhejiang, China
{jiajiong, yingming, xilizju, zhongfei}@zju.edu.cn
[†]Phone: +86-15957112753

## Abstract

In this paper, we study the problem of facial attribute learning. In particular, we propose a Face Recognition guided facial Attribute classification Network, called FR-ANet. All the attributes share low-level features, while high-level features are specially learned for attribute groups. Further, to utilize the identity information, high-level features are merged to perform face identity recognition. The experimental results on CelebA and LFWA datasets demonstrate the promise of the FR-ANet.

## Introduction

Facial attribute classification has many applications including face identification and retrieval. However, facial attribute classification is still very challenging because of large illumination, pose and expression changes. To improve the performance of attribute classification, some efforts have been made to utilize identity information. For example, (Zhong, Sullivan, and Li 2016) pretrains a face identification network to extract features as the inputs of attribute classifiers. However, since the networks are trained as identity classifiers, the performances on identity-free attributes such as Wearing Necktie, Smiling and Wavy Hair are harmed.

On the other hand, modeling attribute correlations is essential for high performance. For instance, (Hand and Chellappa 2016) designs a Multi-task CNN (MCNN) with a grouping scheme. In particular, MCNN shares low-level layers for all the 40 attributes before splitting the network into 6 branches, each one of which focuses on a certain attribute group. By modeling attribute correlations, MCNN has been the state-of-the-art on this task. However, since the statistics characteristics of attributes are rather consistent for the same identity, modeling the relationship between identity and attributes can further improve the performance, which is ignored by MCNN.

We propose a Face Recognition guided facial Attribute classification neural Network, so called FR-ANet, which exploits both attribute correlations and identity information. In particular, a 10-layer-depth network is proposed to share its lower layers for all the 40 attributes while being split into several groups for modeling attribute correlations. Further, the high-level features are merged at a certain depth as the input of an auxiliary face identity recognition task. The face identity recognition task efficiently guides the network to model identity-attribute relationship, since it forces the network to make similar attribute predictions for the same person. The performance of FR-ANet is significantly better than the state-of-the-art methods according to the experimental results on CelebA and LFWA.

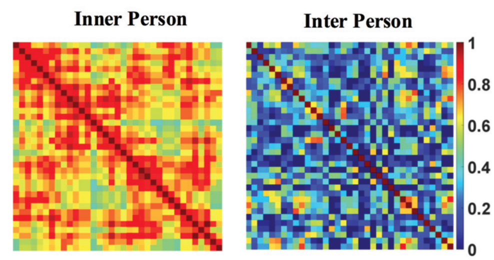


Figure 1: Heat maps for attribute correlations of 35 images within the same person (left column) and among different identities (right column).

## FR-ANet Architecture

We investigate facial attribute classification under a multi-task learning scenario. Since multi-task learning assumes that related tasks should be correlated via a certain structure, we propose to share low-level features for all tasks and to learn high-level features for strongly correlated tasks. It is natural to adopt bifurcation strategy in CNNs, so that all the layers are shared by all the 40 attributes until they are split into several attribute groups and features are specifically learned for each group.

Besides, since the identity information of samples usually exists in the training data, attribute correlations are analyzed on a collection of 35 images within a person and among different identities respectively. As shown in Fig. 1, we find that most images of the same person have similar attribute labels, while attributes for different identities are relatively independent. To model the label consistence for the identity, we develop an auxiliary face identity recognition task

[*]Corresponding author

to regularize the attribute learning. Consequently, FR-ANet is proposed to exploit the attribute and identity information simultaneously to improve the generalization performance of the attribute learning.

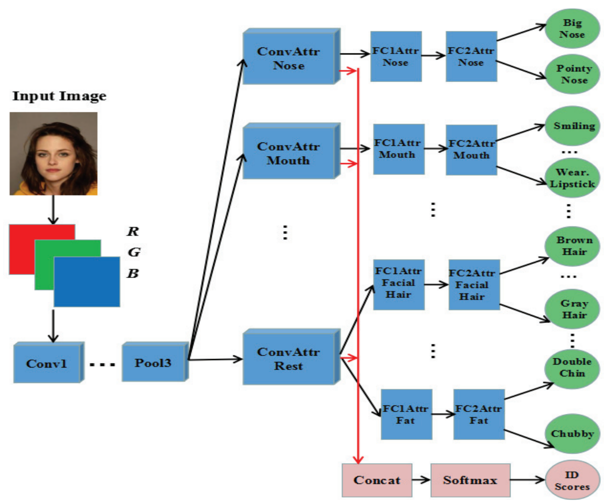


Figure 2: Architecture of FR-ANet.

The FR-ANet architecture is illustrated in Fig. 2, where lower layers are shared by all the 40 attributes until the network is split into 6 groups: Eyes, Nose, Mouth, Face, Gender and Rest. The Rest Group is further split into 4 smaller groups: FacialHair, AroundHead, Cheeks and Fat. And higher layers are specifically learned within groups. Further, feature maps of the 6 groups are concatenated and fed into a face identity recognition task. Therefore, object function of FR-ANet consists of two parts, $Loss_{ATTR}$ and $Loss_{ID}$.

$$Loss_{ATTR} = \sum_{i=1}^{T} loss_i, \tag{1}$$

where $loss_i$ is the classification loss for $i$-th attribute.

$$Loss_{ID} = -log(\hat{p}_k), \tag{2}$$

where $\hat{p}_k$ is the predicted possibility of the $k$-th identity. Then, the object function, $LOSS$, can be formulated as the weighted sum of the two losses:

$$LOSS = Loss_{ATTR} + \lambda \times Loss_{ID}, \tag{3}$$

where $Loss_{ID}$ acts as a regularization term and guide the attribute learning.

The advantage of adopting $Loss_{ID}$ as a regularizer lies on two aspects. First, attributes usually correspond to local facial parts and have simple semantic meanings, while identity is a more abstract concept and relates to the whole face. They are complementary on learning better face representations. Second, attribute classification and face identity recognition are jointly optimized in FR-ANet, so that the identity-attribute relationship is modeled.

An appropriate $\lambda$ helps FR-ANet achieve a balance between attribute classification and face identity recognition. As $\lambda$ becomes too large, attribute features are too similar for the same person to model attribute variety. On the contrary, when $\lambda$ is too small, the face identity recognition task has limited influence on the attribute classification tasks. Note that we choose $\lambda = 1$ for FR-ANet.

## Experiments

To verify the effectiveness of FR-ANet, we compare it with MCNN (Hand and Chellappa 2016), ID Net (Zhong, Sullivan, and Li 2016) and LNets+ANets (Liu et al. 2015). Besides, to verify FR-ANets superiority of exploiting identity information, we also provide another way to utilize identity information during attribute learning called FR+ANet, which treats face identity as another attribute and simply adds a branch after Pool3 for the task. In addition, we report the result of FR-ANet with a zero $\lambda$.

As shown Table 1, FR-ANet significantly outperforms all the other methods by considering both attribute correlations and identity-attribute relationship. In particular, by analyzing the results of FR+ANet and FR-ANet, we conclude that a proper network architecture is essential for taking advantage of the identity information. Since FR+ANet treats identity independently with the attributes, the positive influence from the face identity task is minor and it fails to learn the inner-person statistics characteristics of the attributes. On the contrary, by merging mid-level ConvAttr features, FR-ANet takes full advantage of the identity information. In addition, when we compare the results of FR-ANet and MCNN, the face identity recognition helps improve the attribute predictions performance significantly, which indicates the importance of modeling identity-attribute relationship.

Table 1: Experimental results on CelebA and LFWA.

| Methods | CelebA | LFWA |
|---|---|---|
| LNets+ANet | 87 | 84 |
| ID Net | 88.75 | 84.64 |
| MCNN | 91.25 | 86.27 |
| FR+ANet | 91.28 | 86.30 |
| FR-ANet ($\lambda = 0$) | 91.26 | 86.25 |
| FR-ANet | **92.16** | **87.38** |


## Acknowledgments

This work was supported by National Natural Science Foundation of China (No. 61702448, 61672456) and the Fundamental Research Funds for the Central Universities (No. 2017QNA5008, 2017FZA5007).


## References


Hand, E. M., and Chellappa, R. 2016. Attributes for improved attributes: A multi-task network for attribute classification. *arXiv preprint arXiv:1604.07360*.

Liu, Z.; Luo, P.; Wang, X.; and Tang, X. 2015. Deep learning face attributes in the wild. In *ICCV*, 3730–3738.

Zhong, Y.; Sullivan, J.; and Li, H. 2016. Face attribute prediction using off-the-shelf cnn features. In *Biometrics (ICB), 2016 International Conference on*, 1–7.