Landmark Free Face Attribute Prediction

Geometric Transformation Learning in Face Attribute Prediction

Jianshu Li [1 2] **Fang Zhao** [1] **Jiashi Feng** [1] **Sujoy Roy** [2] **Shuicheng Yan** [1] **Terence Sim** [1]

Abstract

The interplay between geometry and machine learning contains both statistical theory and rich applications. In this paper, we demonstrate an interesting application of geometric transformation learning to the task of face attribute prediction. We show that geometric transformations for facial images can be learned automatically with machine learning approaches, which benefit face attribute prediction tasks[1].

Introduction Face attribute prediction refers to the task of classifying/recognizing attributes, such as facial emotion, gender, age, *etc*, from human face images. Face attribute prediction is usually tackled via a Detection-Alignment-Recognition (DAR) pipeline (Ehrlich et al., 2016). In the DAR pipeline, the Alignment step includes facial landmark detection from landmark detectors (Lai et al., 2016) and hand-crafted geometric transformation estimation from the detected landmarks and some canonical locations. The faces are aligned by the hand-crafted transformation to facilitate the task of face attribute prediction. In this work, we demonstrate that the hand-crafted geometric transformation can be obtained with machine learning, and the learned geometric transformation is more beneficial in the face attribute prediction task compared to the hard-crafted one.

Method In this work, we propose a novel lAndmark Free Face AttrIbute pRediction (AFFAIR) method, which directly learns a hierarchy of geometric transformations for input faces, getting rid of reliance on landmarks and hard-crafted face alignment in DAR. The hierarchy of geometric transformations includes global transformations and part localizations, which employs Spatial Transformer Network (STN) (Jaderberg et al., 2015) and are end-to-end trainable with the attribute prediction task.

The proposed AFFAIR consists of a global TransNet and part LocNets, responsible for learning global geometric transformations and part localizations respectively. The global TransNet takes as input detected faces, and produces a set of optimized geometric transformation parameters $\mathbf{T}_g = \left[\begin{smallmatrix} \theta_{11} & \theta_{12} & d_x \\ \theta_{21} & \theta_{22} & d_y \end{smallmatrix}\right]$ tailored for each input face. The transformation establishes the mapping between the globally transformed face image and the input face image via $[x_i^{\text{input}}, y_i^{\text{input}}]^T = \mathbf{T}_g[x_i^g, y_i^g, 1]^T$. We design a novel competitive learning strategy to ensure the global TransNet learns good global transformations. The part LocNets are designed to capture critical part information for face attribute prediction, since most attributes are reflected by only a small part of the face. The set of part localization parameters is denoted as $\mathbf{T}_p = \left[\begin{smallmatrix} s_x & 0 & d_x \\ 0 & s_y & d_y \end{smallmatrix}\right]$ and the correspondence between the partial face and the globally transformed face image is modelled by $[x_i^g, y_i^g]^T = \mathbf{T}_p[x_i^p, y_i^p, 1]^T$, which transforms the globally transformed face image into face parts. Features are extracted from both the globally transformed face and face parts to perform face attribute prediction.

Figure 1. Left: Learned average face with increasing training epochs. Right: Learned global geometric transformations.

Experiment We conduct experiments on face attribute prediction datasets CelebA (Liu et al., 2015) and MTFL (Zhang et al., 2014). The learned global geometric transformations on faces in CelebA dataset are visualized in Fig. 1. We can see that AFFAIR gradually learns the alignment process in the DAR pipeline during training, and the resultant average face converges to the frontal face. The learned geometric transformations are visualized by grids, which can align input faces with various poses. The learned part localizations are shown in Fig. 2. The regions localized by AFFAIR agree with the actual locations of the attributes.

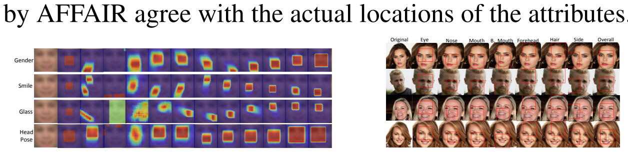

Figure 2. Left: Part localizations on MTFL dataset along different training epochs. We can see the part LocNets search the whole face before converging to the most discriminative parts. Right: Part localization for attributes on CelebA dataset.

Moreover, we find that the learned geometric transformations are more favourable compared to the hand-crafted ones in terms of attribute prediction accuracies, especially when the facial landmarks are difficult to detect.

Conclusion In the paper, we show that geometric learning of faces can be achieved by machine learning, and the learned geometric transformations are not only reasonable, but also beneficial to the face attribute prediction task.

[1]National University of Singapore [2]SAP Innovation Center Networks, Singapore. Correspondence to: Jianshu Li <jianshu@u.nus.edu>.

[1]This is an abstract version of (Li et al., 2018)

References

Ehrlich, Max, Shields, Timothy J, Almaev, Timur, and Amer, Mohamed R. Facial attributes classification using multi-task representation learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 47–55, 2016.

Jaderberg, Max, Simonyan, Karen, Zisserman, Andrew, et al. Spatial transformer networks. In *Advances in Neural Information Processing Systems*, pp. 2017–2025, 2015.

Lai, Hanjiang, Xiao, Shengtao, Pan, Yan, Cui, Zhen, Feng, Jiashi, Xu, Chunyan, Yin, Jian, and Yan, Shuicheng. Deep recurrent regression for facial landmark detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 2016.

Li, Jianshu, Zhao, Fang, Feng, Jiashi, Roy, Sujoy, Yan, Shuicheng, and Sim, Terence. Landmark free face attribute prediction. *IEEE Transactions on Image Processing*, pp. 1–1, 2018. ISSN 1057-7149. doi: 10.1109/TIP.2018.2839521.

Liu, Ziwei, Luo, Ping, Wang, Xiaogang, and Tang, Xiaoou. Deep learning face attributes in the wild. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3730–3738, 2015.

Zhang, Zhanpeng, Luo, Ping, Loy, Chen Change, and Tang, Xiaoou. Facial landmark detection by deep multi-task learning. In *European Conference on Computer Vision*, pp. 94–108. Springer, 2014.