International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 07 | July 2019 www.irjet.net p-ISSN: 2395-0072

Prediction of Facial Attribute Without Landmark Information

Jayashree S. Somani[1], Mrs. V. L. Kolhe[2]

[1]Student, Dept. of Computer Engineering, Dr. D. Y. Patil College of Engineering, Akurdi, Pune, Maharashtra, India
[2]Professor, Dept. of Computer Engineering, Dr. D. Y. Patil College of Engineering, Akurdi, Pune, Maharashtra, India

---

Abstract – *Prediction of the face attributes is very challenging because of complex variation in face. Most of the systems which are used for predicting the attributes of face are unable to give the precision as these methods are depending on the landmark detection and canonical positions. The dependency on landmark detector is unable to give satisfactory results on unconstrained faces with large pose angles, occlusion or blurriness which reduces the performance of attribute prediction. The system explained here use an AFFAIR method that gives the face attribute prediction. By learning a global transformation technique and adaptive part localization technique, system provides the most relevant part for predicting a specific attribute on the face. The AFFAIR learns a good transformation for each input face image directly for attribute prediction with greater accuracy.*

Key Words**: AFFAIR, Landmark Detector, Attribute Prediction, Global Transformation, Part localization.**

1. INTRODUCTION

Describing people depending on their feature points like gender, age, hair style and clothing style is an important problem for many applications in face analysis. Previously Detection-Alignment-Recognition (DAR) method [1] is used for detection of the face attributes with landmark detection from images. As this method depends on the quality of landmark detection, it results reduce the performance of the method as it unable to give the fine result on unconstrained faces. Some of the author uses a pre-trained deep Convolutional Neural Networks (CNN)[4][6][1] for face recognition tasks to obtain global face representation and binary linear SVM classifiers are built on the global face representations to classify face attributes. The previous methods use global methods in which entire object for representation learning and attribute prediction is done without part information and the local method which extract features from relevant regions or parts for attribute prediction.

This study aims to investigate the possibility of optimizing facial landmark detection and alignment which are complicated tasks. Here we are studying AFFAIR method [1] which is an aggregation of global and local methods for attribute prediction. AFFAIR provides an end-to-end learning framework for finding the appropriate transformation. The global transformation is used to detect face and generates transformation parameters tailored for the original input face. Part LocNet is used to focus on the most relevant part of the face for attribute prediction. Finally by integrating both global and local representations, we can predict the facial attributes with no requirement of external landmark points for alignment.

2. LITERATURE SURVEY

Jianshu Li et al. [1] published Landmark Free Face Attribute Prediction in which he proposed the AFFAIR method which learns global transformation and part localization. Then he aggregates both global and local features for robust attribute prediction.

Jianlong Fu et al. [2] published Look Closer to See Better: Recurrent Attention Convolutional Neural Network for Fine-Grained Image Recognition in which he proposed recurrent attention convolutional neural network (RA-CNN) for recognizing the ne grained categories like bird species, etc.

Yang Zhong et al. [3] published Face attribute prediction using off-the-shelf CNN features in which he worked with an alternative way of employing the power of deep representations from CNNs. Combining with conventional face localization techniques, he used the off-the-shelf architecture strained for face recognition to build facial descriptors.

Max Ehrlich et al. [4] published Facial Attributes Classification Using Multi-task Representation Learning in which he proposed a model which learns a shared feature representation that is well suited for multiple attribute classification. Then he learns a joint feature representation which enables interaction between different tasks. For learning this shared feature representation the author has used a Restricted Boltzmann Machine (RBM) based model, enhanced with a factored multi-task component to become Multi-Task Restricted Boltzmann Machine (MT-RBM).

Hamdi Dibekliolu et al. [5] published Combining Facial Dynamics with Appearance for Age Estimation in which he proposed a method which extracts and uses dynamic features for age estimation, using a person's smile.

Kaiming He et al. [6] published Deep Residual Learning for Image Recognition in which he has presented a residual learning framework to ease the training of networks that are substantially deeper than those used previously. The depth of representations is of central importance for many visual

recognition tasks. By doing this, they obtain a 28% relative improvement on the COCO object detection dataset.

Andreas Steger et al. [7] published Failure Detection for Facial Landmark Detectors in which he studied two top recent facial landmark detectors (AFLW, HELEN) and devise confidence models for their outputs. Because of this approach, it correctly identifies more than 40% of the failures in the outputs of the landmark detectors.

Yue Wu et al. [8] published Robust Facial Landmark Detection under Significant Head Poses and Occlusion in which he proposed a unified robust cascade regression framework that can handle both images with severe occlusion and images with large head poses. He introduced a supervised regression method that gradually updates the landmark visibility probabilities in each iteration to achieve robustness.

3. SYSTEM DESIGN

Human face attribute estimation has received a large amount of attention in recent years in visual recognition research because a face attribute provides a wide variety of salient information such as age, gender e. t. c. The fig.1 shows the architecture of the system. The working of the system is as follows:

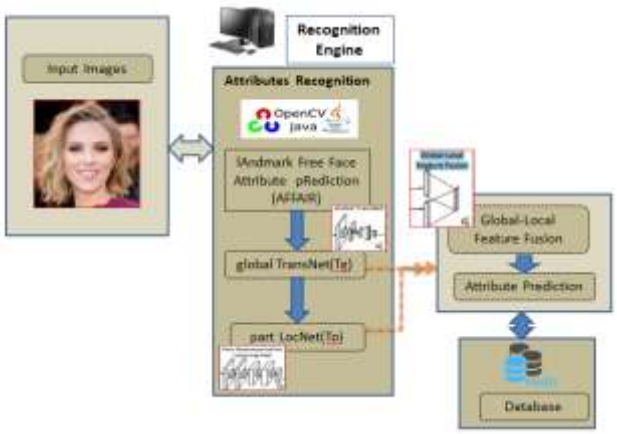

Fig.1: System Architecture

- **Input image:** User uploads images using live camera. Face is detected from the input image and given to the AFFAIR framework as an input. The AFFAIR method is an aggregation of both global and local methods for attribute prediction that integrates both global and local representations and requires no external landmark points for alignment.

- **lAndmark Free Face AttrIbute pRediction method:** lAndmark Free Face AttrIbute pRediction (AFFAIR) method learns a global transformation and part localizations on each input face end-to-end. It mainly has two important components:

✓ *Global Transformation Network:* The global transformation transforms the input face to the one with an optimized configuration for further representation learning and attributes prediction. Global transformation consists of two parts: Global TransNet and Global Representation Learning Net. The global TransNet was followed by global representation learning network which was used to consider all the facial attributes simultaneously. Thus, the global TransNet and the global representation learning net is trained end-to-end for attribute prediction.
The global TransNet in AFFAIR takes the detected face as input, and produces a set of optimized transformation parameters 'T_g' tailored for the original input face for attribute representation learning. The transformation maps the globally transformed face image with the input image via-

$$\begin{pmatrix} x_i^{input} \\ y_i^{input} \end{pmatrix} = T_g \begin{pmatrix} x_i^g \\ y_i^g \\ 1 \end{pmatrix}. \quad (1)$$

Because of this, the globally transformed face images were obtained pixel by pixel. The pixel value at location (x_i^g, y_i^g) of the transformed image was obtained by bilinear interpolating the pixel values on the input face image centered at (x_i^{input}, y_i^{input}).
After this, the globally transformed face image in the pixel format is then given as input to the Global Representation Learning Net which simultaneously considers all the facial attributes. The output face from the global TransNet was denoted by $\theta^T_g(I)$. Then the global face representation learning net, parameterized by θ^F_g, maps the transformed image from the raw pixel space to a feature space beneficial for predicting all the facial attributes. All the facial attributes were denoted by θ^F_g, $\theta^T_g(I)$.

✓ *Part Localization Network:*
Part LocNet was used to localize the most relevant and discriminative parts for a specific attribute and make attribute prediction. LocNet can access the whole face.
The part LocNet predicts a set of localization parameters and it focuses on relevant part on the face through learned scaling and translating transformations. For example, the shape of the eyebrow or the appearance of the goatee, etc. are very small attributes on the face which can be predicted by part localization. The set of part localization parameter was denoted as T_p and the correspondence between the parts to the globally transformed face image was modeled by the following equation which links the pixel value at (x^p_i, y^p_i) on the output partial face image to the pixel

values centered at location (x^g_i, y^g_i) on the globally transformed face image.

$$\begin{pmatrix} x_i^g \\ y_i^g \end{pmatrix} = \mathbf{T}_p \begin{pmatrix} x_i^p \\ y_i^p \\ 1 \end{pmatrix} \quad (2)$$

Part LocNet is also end-to-end trainable. It positions the focus window to a relevant part on the face through learned scaling and translating transformations. Then the locally transformed images were then processed by the Local Representation Learning Net for more than one or all attributes of the face. The additional parameter to generate the transformation, T_{pi}, in the part LocNet for the i^{th} attribute is denoted by θ^T_{pi}. Thus the generated transformation was given by-

$$T_{pi} = \theta^T_{pi}, \theta^F_g, \theta^T_g (I) \quad (3)$$

- **Global-Local Feature Fusion :** In this phase, both global and local representations were integrated by finding a good global transformation to rectify the face scale, location and orientation, and identify the most discriminative part on the face for specific attribute prediction without requiring external landmark points for alignment. The global and local features are generated by the global representation learning net and the part representation learning net which were fused for attribute prediction.
- **Attribute Prediction:** The global and local features were generated by the global representation learning net and the part representation learning net, respectively and were fused to get attribute prediction. Finally after combining the Global-Local features, the specific attribute prediction can be done.

4. RESULTS

Fig.2 gives the experimental results of global transformation and part localization.
The globally transformed images possess 3 × 3 grids on the transformed face boxes. We can see that the two eyes lie in the center grid. The figure 2 demonstrates that the global TransNet is able to generate good global transformations in the sense that the two eyes are centered in the transformed faces images.

Fig.2: Results of Global TransNet[1]

Also, the localization results from the part LocNets for all the 8 facial attribute categories in the CelebA dataset are shown in fig.3. Each column shows a facial attribute category and each row shows one test image, where the original test face images are displayed in the first column. The next columns show the localization results on the globally transformed face images. The boxes indicate the output from the part LocNets. One can see on top of the globally transformed faces, the part LocNet indeed localizes the most discriminative part of the face. For example, the eye region is localized for predicting attributes "Arched Eyebrows", "Bags Under Eyes", "Eyeglasses", etc. The nose region is localized for attributes "Big Nose" and "Pointy Nose".

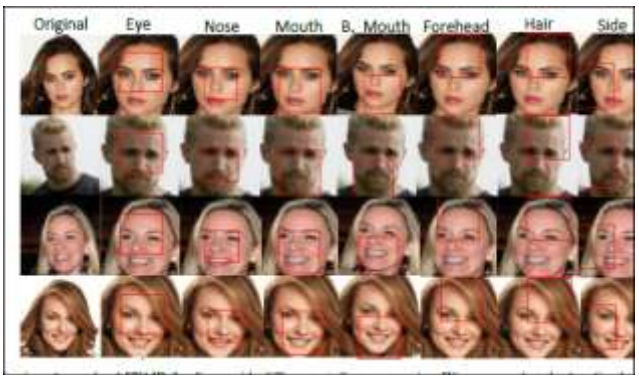

Fig.3: Results of Part localization [1]

5. ADVANTAGES AND DISADVANTAGES

- **Advantages:**
  - AFFAIR model is able to localize the specific part for prediction of the facial attribute with the use of transformation-localization network.
  - The AFFAIR model integrates both global and local representations, by removing the need of external landmark points for alignment.
  - AFFAIR focuses on the local region and learns more discriminative representation for better attribute prediction.
  - AFFAIR does not require face alignment as preprocessing and provides state-of-the-art results for the CelebA, LFWA and MTFL datasets.
- **Disadvantage**:
  - This method gives better performance in all the cases but when the person wears the mask then we cannot detect the face attributes.

6. CONCLUSION

The landmark free Face Attribute prediction (AFFAIR) system is addressed in this paper which does not

depends on landmarks and hardwired face alignment for prediction of the attributes. By learning global transformation, it generates the optimized transformation tailored for each input face. By learning part localization, it locates the most relevant facial part. Finally by aggregating the global and local features attribute prediction is done.

REFERENCES

[1] Jianshu Li, Fang Zhao, Jiashi Feng, Sujoy Roy, Shuicheng Yan, Terrence Sim,"Landmark Free Face Attribute Prediction", IEEE Transactions on Image Processing, Volume: 27, Issue: 9, pages 4651-4662, Sept. 2018.

[2] Jianlong Fu; Heliang Zheng ; Tao Mei, "Look Closer to See Better: Recurrent Attention Convolutional Neural Network for Fine-Grained Image Recognition", 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), November 2017.

[3] Yang Zhong ; Josephine Sullivan; Haibo Li, "Face attribute prediction using off-the-shelf CNN features", 2016 International Conference on Biometrics (ICB), June 2016.

[4] Max Ehrlich; Timothy J. Shields; Timur Almaev; Mohamed R. Amer, "Facial Attributes Classification Using Multi-task Representation Learning", 2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), July 2016.

[5] Hamdi Dibekliolu ; Fares Alnajar ; Albert Ali Salah ; Theo Gevers, "Combining Facial Dynamics With Appearance for Age Estimation ", IEEE Transactions on Image Processing, Volume: 24, Issue: 6, pages 1928 - 1943, June 2015

[6] Kaiming He ; Xiangyu Zhang ; Shaoqing Ren ; Jian Sun, "Deep Residual Learning for Image Recognition", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), December 2016.

[7] Andreas Steger, Radu Timofte, "Failure Detection for Facial Landmark Detectors", Asian Conference on Computer Vision ACCV 2016: Computer Vision ACCV 2016 Workshops pages 361-376.

[8] Yue Wu ; Qiang Ji, "Robust Facial Landmark Detection under Signicant Head Poses and Occlusion", 2015 IEEE International Conference on Computer Vision (ICCV), Dec. 2015