Periocular Recognition Fused with Facial Attributes

Facial soft biometric information extracted from the periocular region of the face (e.g., gender, ethnicity, skin color, hair color, beards, scar, face shape, and so on) is ancillary information to some extend easily distinguished at a distance but it is not fully distinctive by itself in face or periocular recognition tasks. However, the facial attributes can be explicitly fused with periocular recognition algorithms to improve the overall recognition when confronting high variability conditions. We hypothesize that facial attributes can provide valuable information for video surveillance face recognition, where face images are usually captured in poor quality conditions due to variability in distance, illumination and pose. We propose to complement ‘hard’ periocular facial signatures with soft facial attributes such as gender, ethnicity, age, skin color, hair color, face shape and size to improve the overall periocular recognition performance. We assume that using pre-processing techniques (i.e., deblurring and super-resolution), we can improve the resolution and quality of face images such that we can extract reliable facial attributes to complement our primary face recognition algorithm. Although, previous researchers [1]-[4] have tried to use demographic information and soft facial attributes as auxiliary information to improve the performance of biometric systems, but a deep face recognition algorithm that is simultaneously trained and enhanced with a deep facial attribute prediction has not been experimented.

Experimental evidence from our previous work on multimodal face recognition (i.e., soft and hard modalities) [Fariborz CVPRW18 5] have shown that we can improve our CNN-based face recognition model by incorporating soft facial biometric such as age and gender as auxiliary information. Therefore, in this section we propose a new deep face recognition framework which simultaneously predicts facial attributes and incorporates this information with a FR algorithm to identify face images with higher accuracy. Our proposed model is an end to end framework as shown in Fig. 6 which uses several shared convolutional neural network (CNN) layers (common network) and then spreads out into two separate branches (modality dedicated layers); the first branch predicts facial attributes while the second branch identifies face images. Contrary to the existing methods which only use a shared CNN feature space to train these two tasks jointly, we also fuse predicted attribute features with periocular features in the training step to improve the overall face identification performance as shown in Fig. 6. Thus, our proposed deep model jointly predicts facial attributes and identifies people while simultaneously leverages the predicted facial attribute features as an auxiliary modality to improve face identification performance. Our proposed deep model is constructed from two successive cascaded networks as shown in Fig.6. The first common network (net@1) uses the VGG19-like architecture where a global average pooling (GAP) layer is placed after the last convolutional layer. The GAP layer simply takes average of each feature map obtained from last convolutional layer. The second network (net@2) is divided into two separate dedicated branches trained simultaneously while communicating information together through the training process. Both of these two branches consist of two fully connected (FC) layers operating on the output of the first network. The first branch performs attribute prediction task and output of last FC layer in this branch before performing soft-max operation is fused with the GAP layer of net@1 by using a feature fusion technique (i.e. concatenation, sum, average, Kronecker feature products, or more advanced techniques such as multimodal compact bilinear pooling [6], etc.) and finally this fused layer is employed to train the second branch - the face identification task. The overall proposed architecture is shown in Fig. 6, attributes are predicted by features from net@1 and the first branch of net@2 parameters (branch@1) while face images are identified by features from net@1 and all parameters in net@2 forming branch@2. ***It should be pointed that the output of the fusion layer can used as a robust feature vector to be combined with other whole-body feature vectors for analysis.*** We will experiment with the following databases that have annotated soft biometrics: The Labeled Faces in the Wild (LFW) dataset, Celeb-Faces Attributes (CelebA) dataset with 40 facial attributes, PubFig dataset with 72 facial attributes, YouTube Faces video dataset with attribute labels.Point and Shoot Face Recognition Challenge (PaSC) which includes 9,376 still images and 2,802 videos of 293 people, and Soft Biometric Dataset of University of Southampton which includes facial and body attributes.

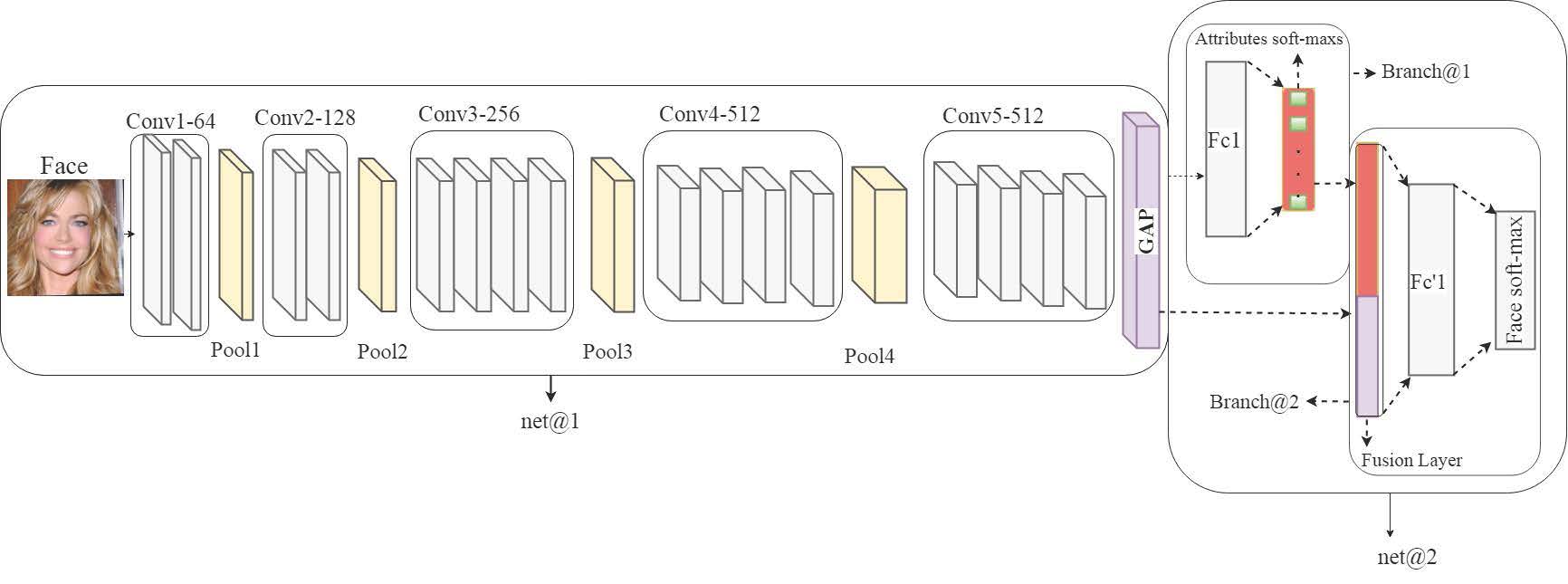


Figure 1: Joint Periocular recognition with facial attributes.

