# AAFACE: Attribute-aware Attentional Network for Face Recognition

Niloufar Alipour Talemi and Nasser M. Nasrabadi

## Introduction

- Despite the considerable progress of face recognition (FR) methods, it is still challenging to recognize faces in unconstrained scenarios due to several factors such as variances in the head pose, inherent sensor noise, and illumination conditions.
- There are several multi-task learning frameworks in biometrics that build synergy among the highly related tasks to boost their individual performances.
- In this work, we present a new multi-branch neural network that simultaneously performs soft biometric (SB) prediction as an auxiliary modality and FR as the main task to improve the performance of our FR model in challenging scenarios.

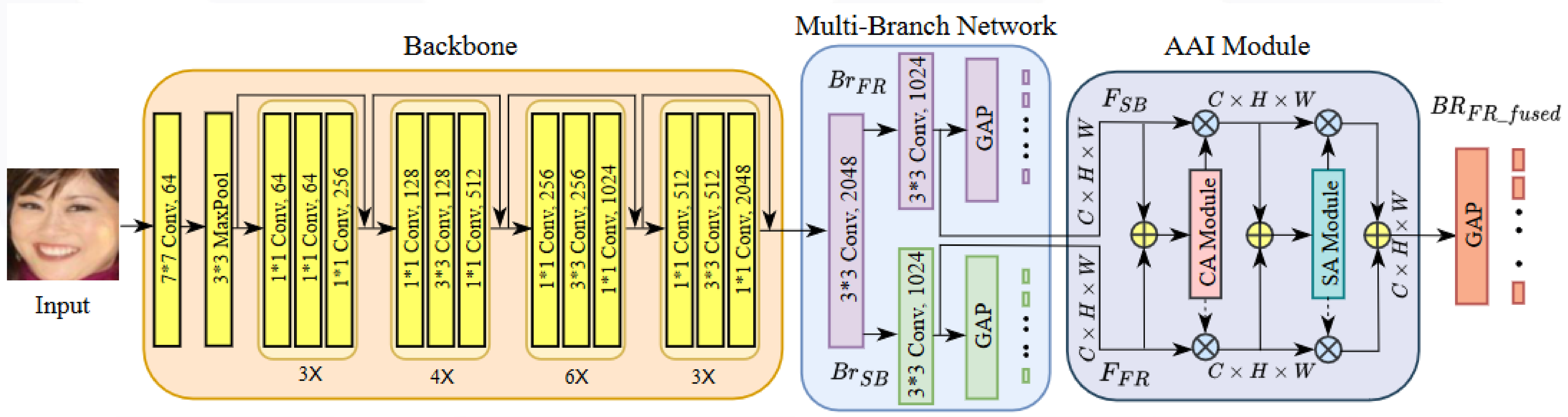
## Objectives

- To utilize facial attributes to enhance the performance of FR, it is important to predict the SB information accurately.
- To efficiently integrate FR features with SB information, we use an **attentional feature-level integration strategy.**

## Proposed Method

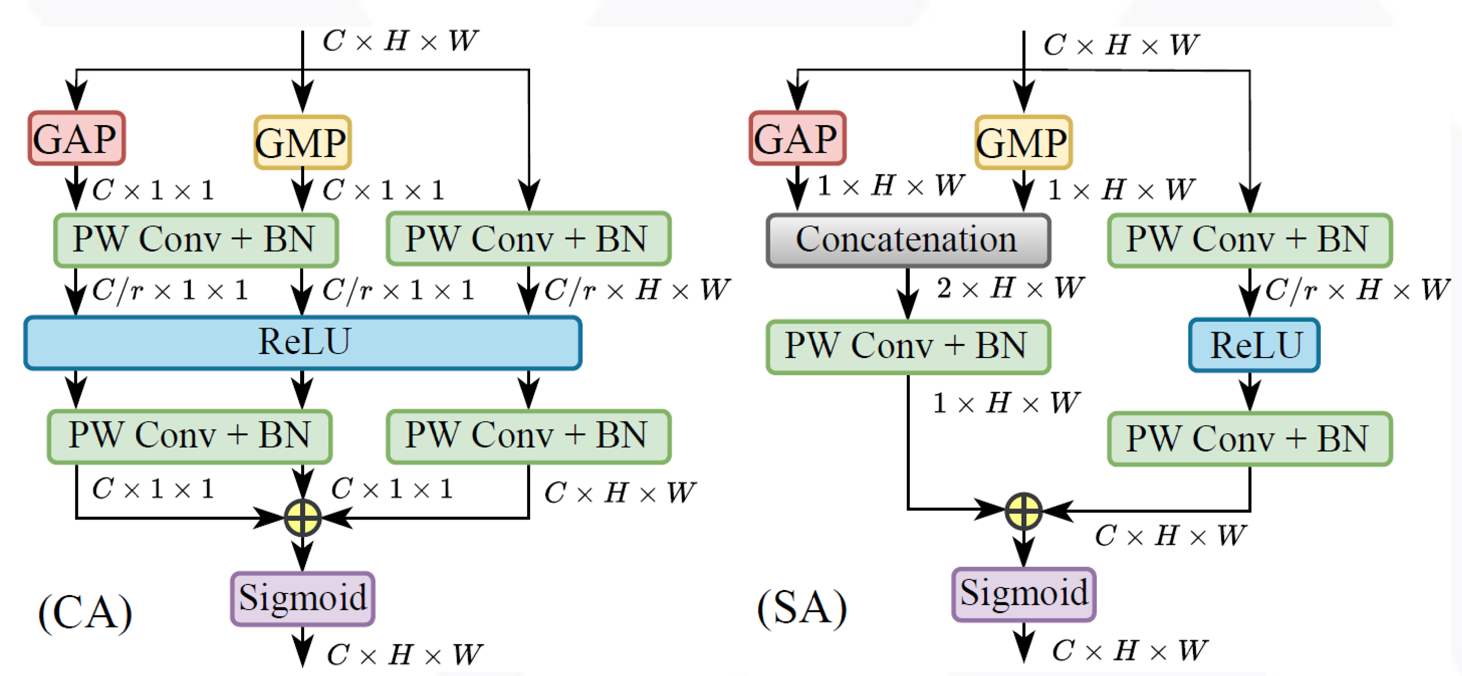
- We propose a multi-branch neural network that simultaneously performs SB prediction and FR in order to enhance the performance of our FR model.
- To effectively leverage SB information for FR, we adopt a feature-level integration strategy through our attribute-aware attentional integration (AAI) module.
- Our context-aware AAI module employs novel multi-scale attention sub-modules to highlight informative features through both channel and spatial dimensions.

## Model




**AAI Module:**

The proposed AAI module has two sequential sub-modules which are channel and spatial attention, respectively. Given two feature maps, $F_{FR}$and $F_{SB}$, the channel-based integration weight, $M_c$, is computed from the multi-scale channel sub-module and then this integration weight will be multiplied by the $F_{FR}$ feature (i.e., $F_{FR} \times M_c$). However, when it comes to the other feature map, $F_{SB}$, the complementary value of the integration weight will be multiplied by the $F_{SB}$ feature (i.e., $F_{SB} \times (1 - M_c)$). Then, the channel-based weighted averaging between $F_{FR}$ and $F_{SB}$ will be given as input to the multi-scale spatial sub-module. Similar to the channel sub-module, spatial-based weighted averaging will be computed between $F_{FR} \times M_s$and $F_{SB} \times (1 - M_s)$.




## Results

Table 1: Classification comparison in terms of accuracy (%) between the proposed SB predictor and the SoTA methods on the CelebA dataset.

| Methods | Bald (B) | Big Nose (BN) | Chubby (CH) | Male (M) | Narrow Eye (NE) |
|---|---|---|---|---|---|
| Z. Liu | 98.00 | 78.00 | 91.00 | 98.00 | 81.00 |
| Moon | 98.77 | 84.00 | 95.44 | 98.10 | 86.52 |
| HyperFace | - | - | - | 97.00 | - |
| R. Ranjan | - | - | - | 99.00 | - |
| MCFA | 99.00 | 84.00 | 96.00 | 98.00 | 87.00 |
| L. Mao | 99.03 | 84.78 | 95.86 | 98.29 | 87.73 |
| Ours | 98.14 | 83.27 | 95.48 | 98.74 | 85.26 |
| Ours (jointly) | **99.10** | **84.84** | **96.09** | **99.16** | 87.56 |

Table 2: Performance comparison of our proposed method (AAFace) with recent SoTA FR methods. TAR is reported at FAR = 0.01%.

| Methods | Verification Accuracy | | | | TAR* | |
|---|---|---|---|---|---|---|
| | LFW | CFP-FP | CPLFW | AgeDB | IJB-B | IJB-C |
| CosFace | 99.81 | 98.12 | 92.28 | 98.11 | 94.80 | 96.37 |
| ArcFace | 99.83 | 98.27 | 92.08 | 98.28 | 94.25 | 96.03 |
| MV-Softmax | 99.80 | 98.28 | 92.83 | 97.95 | 93.60 | 95.20 |
| MagFace | 99.83 | 98.46 | 92.87 | 98.17 | 94.51 | 95.97 |
| SCF-ArcFace | 99.82 | 98.40 | 93.16 | 98.30 | 94.74 | 96.09 |
| AdaFace | 99.82 | 98.49 | 93.53 | 98.05 | 95.67 | 96.89 |
| AAFace | 99.82 | **98.56** | **93.71** | **98.24** | **95.70** | **96.93** |

## Qualitative Evaluation

| | | | | | | |
|---|---|---|---|---|---|---|
| **Bald** | N | Y | N | N | N | N |
| **Big Nose** | N | Y | N | Y | N | N |
| **Chubby** | N | N | N | Y | N | N |
| **Male** | N | Y | N | Y | Y | N |
| **Narrow Eye** | N | N | N | N | Y | N |

## Conclusion

- We proposed a multi-branch neural network that uses shared CNN feature space for two related tasks which are the SB prediction and FR.
- The proposed architecture, not only predicts the SB attributes and simultaneously identifies face images but also utilizes SB attributes as auxiliary information to improve the performance of our FR model.
- Results demonstrate that training both tasks jointly improves their performance in comparison with separate training.

## Key References


1. Ajeev Ranjan, Swami Sankaranarayanan, Carlos D Castillo, and Rama Chellappa, "An all-in-one convolutional neural network for face analysis," in FG, 2017, pp. 17–24.
2. Minchul Kim, Anil K Jain, and Xiaoming Liu, "Adaface: Quality adaptive margin for face recognition," in CVPR, 2022, pp. 18750–18759.
3. Anghyun Woo, Jongchan Park, Joon-Young Lee, and In SoKweon, "CBAM: Convolutional block attention module," in ECCV, September 2018.


## Acknowledgements

This research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA R&D Contract No. 2022-21102100001.

West Virginia University
Lane Department of Computer Science and Electrical Engineering

Funded by Intelligence Advanced Research Projects Activity (IARPA)