SAF- BAGE: Salient Approach for Facial Soft-Biometric Classification - Age, Gender, and Facial Expression

Ayesha Gurnani Kenil Shah Vandit Gajjar Viraj Mavani
Yash Khandhediya

IEEE 2019 Winter Conference on Applications of Computer Vision

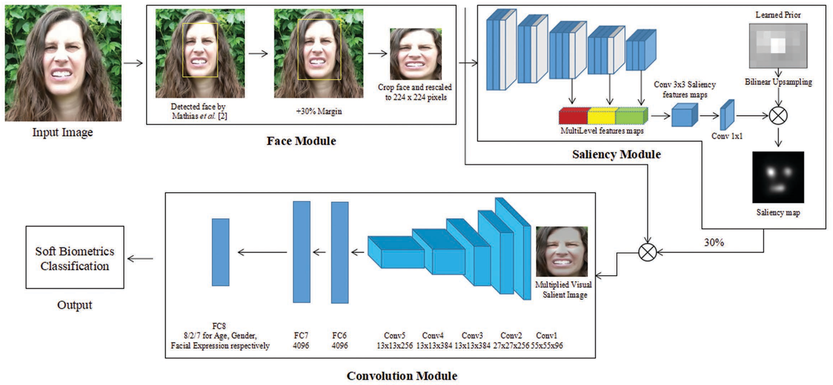
Motivation

- Background in images pose a problem in increasing the intra - class variance and thereby also hinders the classification performance to some extent.
- Eliminating or reducing the effect of background data would contribute in increasing the overall classification performance of any model.

Idea

- Visual Saliency or Attention maps provide the regions of an image where in the subject of interest is located in different forms.
- We believe that the intensity maps computed using the cognitive theory of visual saliency would contribute in localization of the subject in the frame and can be useful to reduce the impact of background on intra-class variance.

Our Approach



Results

Method	Accuracy
Eidinger *et al.*	45.1 ± 2.6 %
Levi *et al.* (Single Crop)	49.5 ± 4.4 %
Levi *et al.* (Multi Crop)	50.7 ± 5.1 %
Ours (Without *saliency* Module)	**52.2 ± 3.9 %**
Qawaqneh *et al.*	59.9 %
Dehghan *et al.*	61.3 ± 3.7 %
Ours (With *saliency* Module)	**62.11 ± 3.2%**

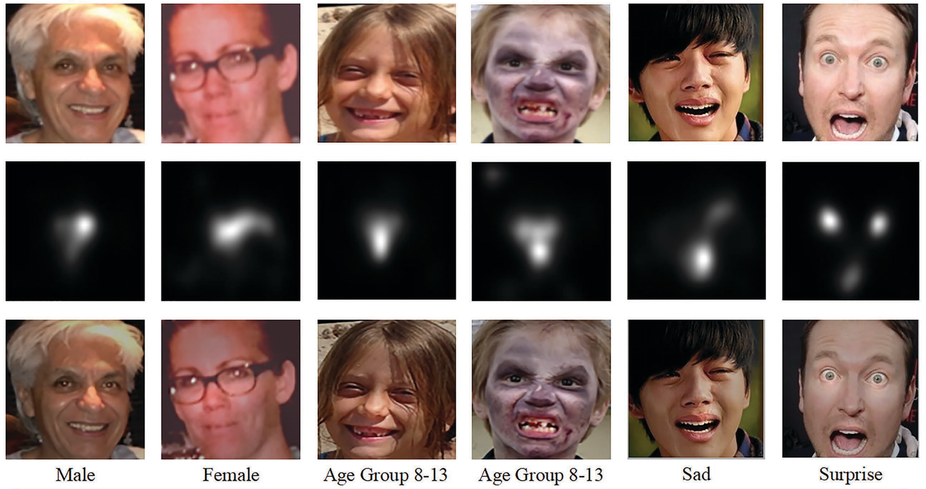
Method	Accuracy
Eidinger *et al.*	77.8 ± 1.3 %
Liao *et al.*	78.63 %
Hassner *et al.*	79.3 ± 0.0 %
Ours (without *saliency* module)	**83.4 ± 1.6 %**
Levi *et al.*	85.9 ± 1.4 %
Levi *et al.*	86.8 ± 1.4 %
Dehghan *et al.*	91 %
Ours (with *saliency* module)	**91.8 ± 1.2 %**

Method	Accuracy
Söderberg *et al.*	48.58 %
Hewitt *et al.*	57.8 %
Ours (without *saliency* module)	**59.2 %**
Mollahosseini *et al.*	64.53 %
Ours (with *saliency* module)	**67.65 %**

Age Group	0-2	4-6	8-13	15-20	25-32	38-43	48-53	60-
0-2	**92.1**	4.2	0.0	1.3	0.0	2.4	0.0	0.0
4-6	22.2	**70.2**	5.8	3.1	0.0	0.0	0.4	0.1
8-13	3.6	12.6	**52.8**	13.2	11.7	3.8	2.3	0.0
15-20	1.4	0.3	13.7	**36.3**	42.6	3.7	1.2	0.8
25-32	0.2	0.0	0.1	4.7	**88.4**	3.8	2.3	0.5
38-43	0.0	0.4	0.7	3.8	49.8	**29.2**	13.8	2.3
48-53	0.8	0.0	0.0	0.9	2.6	18.3	**47.8**	29.6
60-	0.0	0.3	0.4	2.8	1.7	3.5	11.2	**80.1**

	Neutral	Happy	Sad	Surprise	Fear	Disgust	Anger
Neutral	**62.4 (312)**	6.6 (33)	7.8 (39)	7.4 (37)	3.2 (16)	3.8 (19)	8.8 (44)
Happy	3.8 (19)	**76.4 (382)**	2.0 (10)	5.0 (25)	1.2 (6)	8.4 (42)	3.2 (16)
Sad	21.2 (106)	2.8 (14)	**64.4 (322)**	3.4 (17)	4.0 (20)	3.6 (18)	0.6 (3)
Surprise	7.8 (39)	5.0 (25)	3.0 (15)	**73.0 (365)**	8.6 (43)	1.4 (7)	1.2 (6)
Fear	2.2 (11)	5.2 (26)	6.0 (30)	11.6 (58)	**68.4 (342)**	4.8 (24)	1.8 (9)
Disgust	5.0 (25)	3.4 (17)	1.4 (7)	3.8 (19)	18.2 (91)	**66.2 (331)**	2.0 (10)
Anger	6.0 (30)	1.6 (8)	5.8 (29)	5.6 (28)	5.8 (29)	12.4 (62)	**62.8 (314)**

(a) (b) (c) (d) (e) (f)



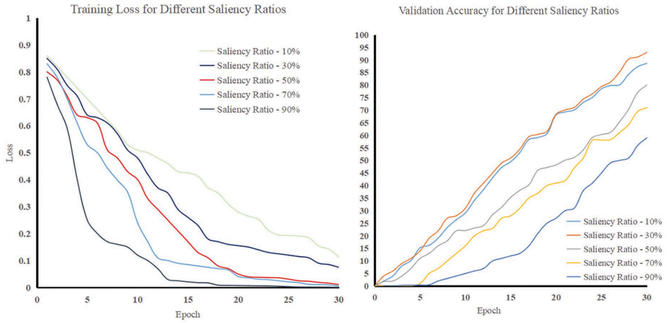
Male Female Age Group 8-13 Age Group 8-13 Sad Surprise

References

[1] Levi, Gil, and Tal Hassner. "Age and gender classification using convolutional neural networks." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2015.
[2] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556. 2014 Sep 4.
[3] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. InAdvances in neural information processing systems 2012 (pp. 1097-1105).
[4] Mollahosseini A, Hasani B, Mahoor M H. Affectnet: A database for facial expression, valence, and arousal computing in the wild. arXiv preprint arXiv:1708.03985. 2017 Aug 14.

Ablation Experiments

- The VGG - 16 network is about 1.6% better than the AlexNet, while the GilNet shows a comparable result with slightly less accuracy of 89.8 ± 1.3%.
- We try different saliency ratios to check the performance of our approach. Therefore, we fine-tuned the AlexNet network with 10%, 50%, 70%, and 90% multiplied reweighted salient face images.



Class Activaton Maps

- Using CAM, we verify the fact that the naturally salient regions of the face i.e. Eyes, Nose and Mouth are also dominant features for classification.