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

Ayesha Gurnani[1,2], Kenil Shah[2], Vandit Gajjar[2,3], Viraj Mavani[1,2], Yash Khandhediya[2,3]

[1]Erik Jonsson School of Engineering and Computer Science, The University of Texas at Dallas, USA
[2]Computer Vision Group, L. D. College of Engineering, India
[3]School of Engineering and Applied Science (SEAS), Ahmedabad University, India
[4]Dosepack LLC, Meditab Software Inc., India

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

- So, we multiply the weighted actual images with the visual saliency intensity maps.
- By this, we get an image where in the subject is a bit more exposed than the background.
- We have conducted experiments supporting the claim that using such images improve the classification performance.

Original Image

Saliency Map

Salient Windowed Image

Experimentation I – Age and Gender

- Age

Method	Accuracy
Eidinger *et al.* [28]	45.1 ± 2.6 %
Levi *et al.* [27] (Single Crop)	49.5 ± 4.4 %
Levi *et al.* [27] (Multi Crop)	50.7 ± 5.1 %
Ours (Without *saliency* Module)	**52.2 ± 3.9 %**
Qawaqneh *et al.* [37]	59.9 %
Dehghan *et al.* [33]	61.3 ± 3.7 %
Ours (With *saliency* Module)	**62.11 ± 3.2%**

- Gender

Method	Accuracy
Eidinger *et al.* [28]	77.8 ± 1.3 %
Liao *et al.* [32]	78.63 %
Hassner *et al.* [38]	79.3 ± 0.0 %
Ours (without *saliency* module)	**83.4 ± 1.6 %**
Levi *et al.* [27]	85.9 ± 1.4 %
Levi *et al.* [27]	86.8 ± 1.4 %
Dehghan *et al.* [33]	91 %
Ours (with *saliency* module)	**91.8 ± 1.2 %**

Experimentation II – Facial Expression (FER)

- AffectNet Benchmark
- The training image size is very high for Facial Expression dataset, thus use of data augmentation is not needed for our fine-tuning process.
- The fine-tuning is accomplished by the last 4 layers {Conv5, fc6, fc7, and fc8} of AlexNet.
- The model is fine-tuned for 40 epochs. The learning rate is set to 0.001 and divided by 10 after 30 epochs. The effective training batch size is set to 128 and dropout set to be 0.35.

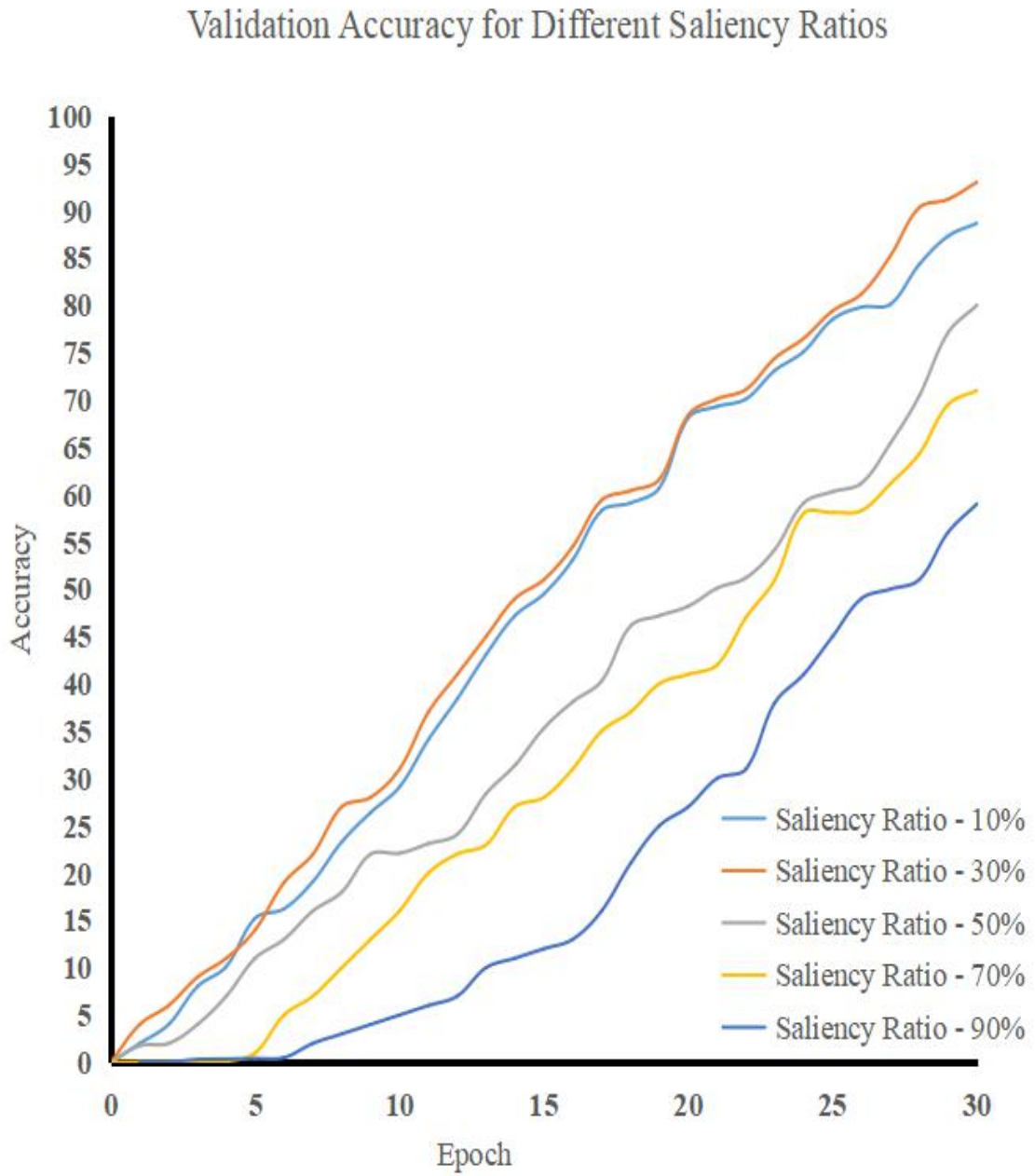
Method	Accuracy
Söderberg *et al.* [40]	48.58 %
Hewitt *et al.* [39]	57.8 %
Ours (without *saliency* module)	**59.2 %**
Mollahosseini *et al.* [36]	64.53 %
Ours (with *saliency* module)	**67.65 %**

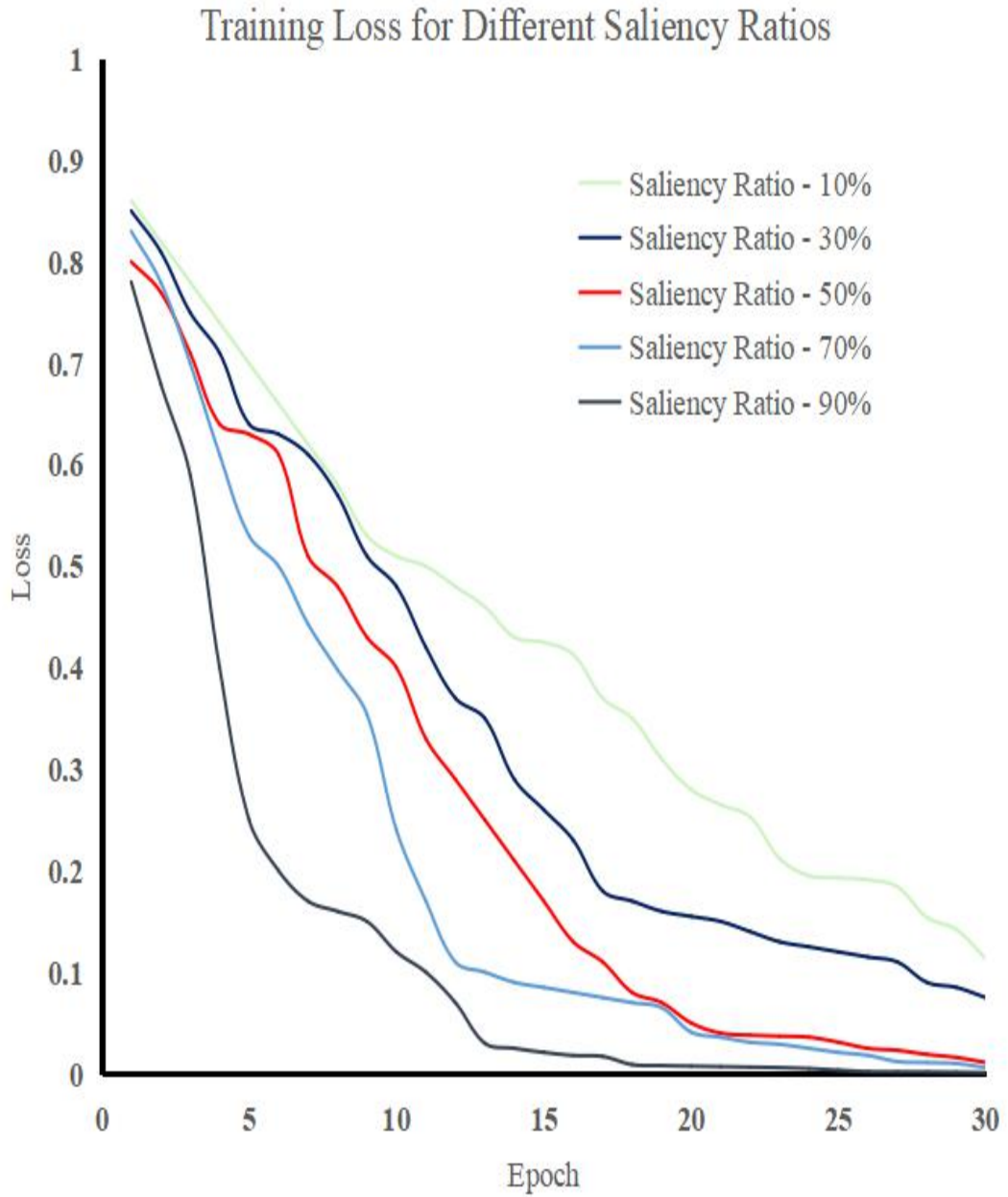
Ablation Study I – Choice of Classification Network

- To test the influence of classification network on the proposed approach, we fine-tune VGG-16 network [6] (Only last fully-connected layer due to computing power) and GilNet [27].
- 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%.

Ablation Study II – Different Multiplication Weighing Ratios

- In our approach, one changeable parameter is the use of 30% reweighted saliency ratios.
- 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.


Validation Accuracy for Different Saliency Ratios
Accuracy
Epoch
Saliency Ratio - 10%
Saliency Ratio - 30%
Saliency Ratio - 50%
Saliency Ratio - 70%
Saliency Ratio - 90%


Training Loss for Different Saliency Ratios
Loss
Epoch
Saliency Ratio - 10%
Saliency Ratio - 30%
Saliency Ratio - 50%
Saliency Ratio - 70%
Saliency Ratio - 90%

Class Activation Maps

- From the salient multiplied images, it can be seen that the salient regions - Eyes, Nose, and Mouth are the dominant attributes which help the model to classify facial soft-biometrics.
- Using Class Activation Maps (CAM), we also verify that these dominant regions are playing a significant role for classification. In CAM, for AlexNet, we remove the fully-connected layers before the final output and replace it with Global Average Pooling (GAP) followed by a softmax layer. We can identify the importance of the image regions by projecting back the weights of the output layer on the feature maps, by giving this simple structure.

Conclusion

- Visual Saliency maps can help improve classification performance by enhancing the focus on the more visually dominant areas of the image.
- This can also be applied to improve the detection task and is open for future extension.

Thank You!