Viability of implementing AFIF Deep Learning algorithm on mobile apps

Nikolay Prieto Ph.D(c)

Reconoser ID Incorporated
Email: enprietop@unal.edu.co

From the literature it can be depicted a variety of algorithms to perform gender classification. However, many of them outperform implementing on generalized datasets or, even further difficult scenarios such as occluded or badly illuminated samples. In order to improve the accuracy on the mentioned scenarios, a proposition made by [1] involves a score fusion based on four specific facial features and the holistic feature of the image i.e the foggy face. Those individual classifications are evaluated by the AdaBoost algorithm to fuse the classification and make a final judgement of gender.
In this report, we are going to evaluate the viability according the technical requirements, the data available to perform this operation, discuss about the pre-processing method and, evaluate the implementation of the Adaboost algorithm.

Keywords: Deep Learning, Gender detection, AdaBoost.

Received August 23rd 2021; revised August 24rd 2021

1. INTRODUCTION

1.1. The SoF dataset

We will briefly describe the AFIF algorithm, from its conception through its performance. Broadly speaking, based on occlusion problems and illumination changes they performed the SoF dataset, *"The SoF comprises* $2,662$ *original images of size* 640×480 *pixels for* 112 *persons (*66 *males and* 46 *females) from different ages. The glasses are the common natural occlusion in all images of the dataset."*

The first part of the dataset was dedicated to obtain pictures at different orientations (± 35) either indoors or outdoors. The second part was dedicated to challenging illumination scenarios where the frontal picture is taken with different lamp angles. The metadata is provided to get more insights from this dataset. A more amplified dataset was generated with noisy filters and translation effects in order to enhance the performance of the DNN architecture.

1.2. The Deep Neural Network

Four independent DNN were applied to each of the most important facial features and one for the foggy face. In order to detect those features, the procedure applied to each image was the following: i) The illumination invariant image filter (I_{SSR}); ii) They selected the facial landmarks; iii) they took the eyes, mouth and nose individual sections; and iv) through a Gaussian filter they generate the holistic feature.

A deep CNN was applied to each feature patches of

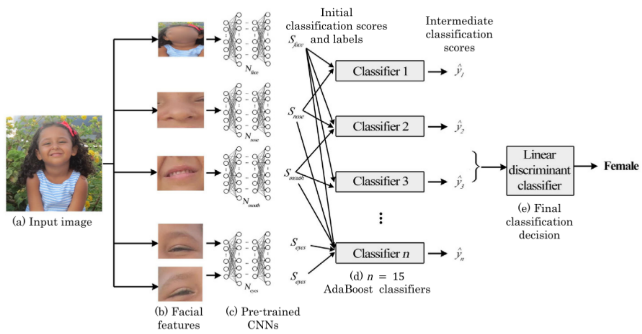

FIGURE 1. General architecture of the AFIF model. Taken from [1]

size Boost classifiers, *"those are trained thereafter on 60% of the rest of the* 227×227 *pixels, extracted from* 75% *of the training set. The Ada–training set using the prediction scores of the pre-trained CNNs. At the final stage of training, the fusion classifier is trained using the estimated classes reported by the AdaBoost classifiers over the rest of the training set. Eventually, we test the entire algorithm using the testing fold."* Fig. 1 depicts the general DNN architecture for all the entire process.

2. IMPLEMENTATION ANALYSIS ON MOBILES APPS.

As the problem mentions, It is understood the mobile applications will rely on cloud computing with sensing from mobile (camera). In this sense, we will divide the viability of the implementation on mobiles based on the performance and capacity, the data collection analysis, the online experimentation and, model debugging and testing.

2.1. Performance and Capacity considerations

Major performance and capacity discussions come in during the following two phases of building a machine learning system:

- **Training time:** How much training data and capacity is needed to build our predictor?
- **Evaluation time:** What are the Service level agreement(SLA) that we have to meet while serving the model and capacity needs?

For this specific case, the training time complexity for Neural Networks is exponential per hidden layer, approximately. As there are three layers per each facial feature, the complexity is $O(4n^3)$ as there are four CNN implied. For the Adaboost algorithm, the literature mentions that the computational complexity is approximated to $O(n\log(n))$ [2].

On the other hand, the evaluation time with CNN is mentioned in the Eq. 1

$$O(fn_{l_1} + n_{l_1}n_{l_2} + \ldots) \tag{1}$$

Where f correspond to the number of features, and n_{l_i} is the number of neurons at the t^{th} layer in a neural network.

Relatively deep neural network takes a lot more time in both training and evaluation. Its need for training data is also high. However, its ability to learn this complex tasks is highly accurate, and it gives more accurate predictions in comparison to other models. Therefore we will need their implementation considerations in the large-scale system.

Once the SLA are established we can analyse further the performance and capacity of the system to determine desired time for the task.

2.2. Data collection analysis:

In the introduction we simplified the description of the data and, in this section we will analyse the implementation of the dataset for a local implementation. This SoF dataset is focused on the occluded photos and different face orientation. However, it is not discriminant to the human race region. According to [3]: *existing public face datasets are strongly biased toward Caucasian faces, and other races (e.g., Latino) are significantly underrepresented. This can lead to inconsistent model accuracy, limit the applicability of face analytic systems to non-White race groups, and adversely affect research findings based on such skewed data.*

In our local application, the accurate model evaluation can be reduced as the data is not balanced

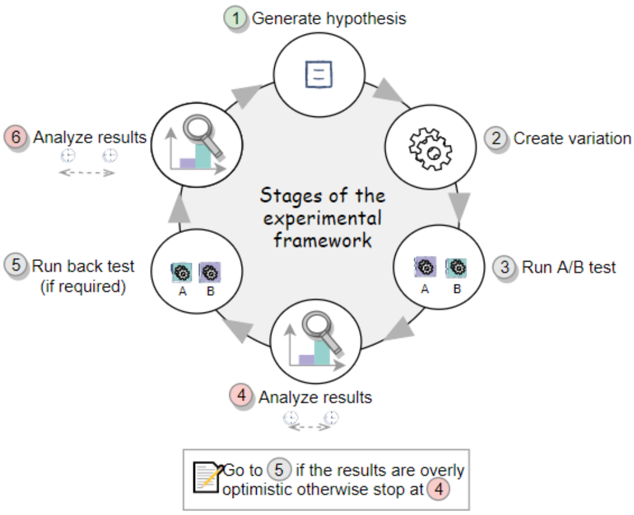

FIGURE 2. General long-term A/B testing procedure. Taken from Educative.io

respect to human races. In this case, a transfer learning process might be implemented with the FareFace dataset in order to increase the accuracy.

2.3. The online Experimentation

A/B testing is very beneficial for gauging the impact of new features or changes in the system on the user experience. It is a method of comparing two versions the app against each other simultaneously to determine which one performs better. A control version with the AFIF proposition pre-trained with the SoF dataset will remain as the control version. On the other hand, a version with the enhanced dataset and some other samples from mobiles can be tested. With statistical hypothesis we can determine if there is any significant difference between version and control. In the long-term scale we can implement the procedure mentioned in the Fig. 2

2.4. Model Debugging and Testing

There are two main phases in terms of the development of a model that we will go over:

- Building the first version of the model and the DL system.
- Iterative improvements on top of the first version as well as debugging issues in large scale DL systems.

The workflow is proposed to deploy the first version and is shown in Fig. 3

It is supposed that train input images keep the same features as mobile samples. However, this might not be true as the way we generated features for our online system might not exactly be the same. It is a common practice to append features offline to our training data for offline training and then add them later to the online

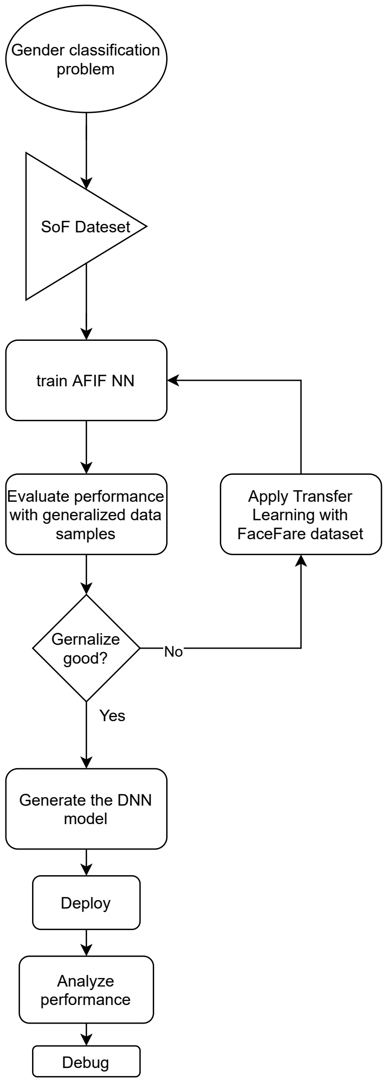

FIGURE 3. Proposed workflow to implement this process on a mobile app.

model serving part. So, if the model does not perform as well as we anticipated online, it would be good to see if feature generation logic is the same for offline training as well as online serving part of model evaluation.

Analyzing the performance, we should take into account the over-fitting problem, as ANN tend to memorize the training data and consequently, outperform the generalize data. Cross-validation techniques should help to overcome this issue. Finally, the debugging process begins with the identification of the problem in order to find the error location. Then we analyze the error and generate ideas for possible solutions.

3. CONCLUSIONS

In this paper we have considered auctions to solve the questions for the interview:

- *Propose the workflow to achieve a successful solution:* The workflow is shown in Fig. 3. We made a general framework in which we provide a local solution on a mobile app.
- *Do you know the dataset for the solution?* In the introduction we briefly described the motivation and how is composed the dataset.
- *Are they relevant?, Why?* This is extremely relevant as it considers different angles in facial image, they also include features from the holistic face shape, the metadata is very completed and, performs well on non-illuminated regions.
- *Give your opinion about the augmenting data process:* Providing Gaussian smoothing, Gaussian noise, posterization filter, nose occlusion, and mouth occlusion filters in the augmented data will help to perform better the evaluation of outdoor and indoor images. Also, small traslations are a good idea which can be implemented easily with libraries as tensorflow (keras), pytorch, or MXNet.
- *Is the pre-processing enough?* For the purpose and the results depicted in the paper's figure those filters performs very well on shadows. Also they detect the facial regions easily. Another idea come up for this process implementing ANN for facial gesture detection. However, it increases the computation cost, thus I would keep this process as long as I play with It.
- *Are there another options for the final classification with AdaBoost?* Most of the tree-based algorithms such as Random Forest, Decision Tree, AdaBoost, etc. have a computational complexity of $O(n\log(n))$. Even though Adaboost shown very good performance, from the literature I would implement Extreme Gradient Boosting algorithm, I have worked previously with it and performs extremely good.

REFERENCES

[1] Afifi, M. and Abdelhamed, A. AFIF4: Deep gender classification based on AdaBoost-based fusion of isolated facial features and foggy faces. *Journal of Visual Communication and Image Representation*, **62**, 77–86.

[2] machine learning - What is the O() runtime complexity of AdaBoost? - Stack Overflow.

[3] Kärkkäinen, K. and Joo, J. FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age. , **?**