OPT 2 – COMPUTER VISION (2022 / UB)
REPORT: **Task 2**

Group members:
Xavier de Juan Pulido, xdejuapu7@alumnes.ub.edu, xavidejuan
Lorenzo Andrés Vigo del Rosso, lvigodel27@alumnes.ub.edu, lorenzovigo

1. SUMMARY OF CONTRIBUTIONS

The strategy followed in this task is based on the provided baseline notebook and the modification of the loss function used in order to work with the imbalanced dataset in terms of gender, age, ethnicity and face expression. To address this problem different strategies have been defined by using different models, training strategies and considering different attributes of the data. Our experiments have been mainly focused on tuning the model architecture and the training strategy. Regarding the model architecture, we used two different pretrained backbones: ResNet50 and InceptionResNetV2, and we realized that ResNet50 was giving better results because the dataset where it was pretrained helped better to address our problem. Then, in the top of this backbone we added some layers to refine the predictions. With respect to the training strategy we tried different approaches: one single step, two steps and even three steps. On the one step training the whole model is trained, on the two step one the backbone is freezed at first step and then the whole model is trained and in the three step process we first freeze the backbone, secondly the head is freezed and at the third step the whole model is trained. Finally, in order to address the dataset imbalance we considered all group combinations in terms of the metadata we had, for example: male, caucassian, 0-20 years, happy. Subsequently we computed a class weight for each one and lastly these weights were used in each training to modify the loss function.

2. EXPERIMENTAL SETUP & DISCUSSION OF THE RESULTS

The setup to carry out will consist of an iPython notebook in which the code runs with Tensorflow GPU v2.4.0. It will be hosted in Google Colab to take advantage of the usage of GPU processing, which allows the executions to be faster.

We are basing our code on a provided baseline notebook. The dataset loading, the pretrained model downloading, model adaption, bias analysis and metrics evaluation have been reused. Then, different backbones, training strategies, top classification layers and hyperparameters have been used in the experiments.
In addition, the goal of this task is to modify the loss function in order to achieve better results and combat the bias discrimination between different groups in terms of gender, age, ethnicity and expression. For this purpose, the dataset samples have been grouped in terms of the previous characteristics resulting in 78 different groups. For each of these groups we assigned a class weight being the total number of samples divided by two times the number of samples of that group. This way, the groups with less samples will be assigned a greater weight when evaluating the loss function.
The following subsections will describe the experiments run. For the experiments the training set will be used for the training step, then the validation set will be used to decide which strategies, hyperparameters or architecture are the best but finally the results reported here

will be evaluated in the test set. Besides, each experiment will follow the class weight approach assigning the class weights as described in the previous paragraph.

EXPERIMENT 1.
In this first experiment the model we used consists of a pretrained ResNet50 on faces as backbone and one classification layer at the top with a sigmoid as activation. The training strategy includes only one step training the whole model using Adam as optimizer with learning rate 1e-5, Mean Squared Error as loss and Mean Absolute error as metric. The batch size has been set to 32, the maximum number of epochs to 50 using an Early Stopping with patience 5 looking to the validation loss.

EXPERIMENT 2.
As a second experiment we used the same strategy and hyperparameters as the previous one but changing the backbone in the model architecture, we used InceptionResNetV2 from keras pretrained on the Imagenet dataset.

EXPERIMENT 3.
This third experiment is similar to the first one, but modifying the top layers from the model architecture. Again, we have used the ResNet50 pretrained on faces as backbone and then we have added a dropout layer with rate 0.5, a fully connected layer with 128 units with ReLu, another fully connected layer with 32 units with ReLu also and finally the classification layer with sigmoid as activation function. The training strategy used in this experiment is exactly the same as the first experiment, one single step training the whole model by using the same hyperparameters.

EXPERIMENT 4.
This experiment is almost the same as the previous one but modifying the backbone in the model architecture, again we have used the InceptionResNetV2 pretrained on imagenet. The training strategy and hyperparameters have been the same used in experiment 3.

EXPERIMENT 5.
This experiment adopted the same model architecture as the third experiment, ResNet50 as backbone with dropout and two fully connected layers at the top. However, the training strategy changes by using two training steps. At the first step, the backbone is freezed and the model is trained with Adam as optimizer with learning rate 1e-4, Mean Squared Error as loss, Mean Absolute Error as metric, 32 as batch size during 50 epochs with an Early Stopping equal to the previous experiments. Then, at the second step, the whole model is trained and with respect to the previous step, the only change is the learning rate used in the Adam optimizer that, in this case, is equal to 1e-5.

EXPERIMENT 6.
The sixth experiment adopts the same training strategy as the previous experiment. We have used the same hyperparameters and top layers the only change on the experiment is the backbone that, again, the InceptionResNetV2 pretrained on Imagenet is used.

EXPERIMENT 7.
In this experiment, we wanted to refactor again the training strategy by including 3 different steps of training. The model architecture is the same as in experiment 5, ResNet50 and the

same top layers. In the first step of training the backbone is freezed and trained the same way as in experiment 5. The second step unfreezes the backbone and freezes the top layers and the only difference with respect to the previous step is the Adam learning rate being 1e-5. Finally, in the third step, the whole model is trained with the same hyperparameters as in the second step.

EXPERIMENT 8.
This experiment is almost equal to the previous one but, as done in the previous experiments, we have changed the backbone with the InceptionResNetV2 pretrained on Imagenet. Then, the training strategy used is the one described in the seventh experiment, the three steps of training.

RESULTS.

Green and red cells show the best and worst results respectively within each kind of evaluation. The less intense shades of these colors show close competitors to the best/worst results.

Experiment	MAE	Age Bias	Gender Bias	Ethnicity Bias	Face Expression Bias
1	30.8841	19.8232	2.0060	3.9053	2.0783
2	14.7199	15.1689	2.2678	2.4592	1.9722
3	14.6654	4.8177	3.3865	3.5401	1.4825
4	14.7797	15.0066	2.0377	2.5554	1.4152
5	9.0614	5.6299	0.3697	0.8153	0.3491
6	15.3635	15.3791	1.8984	2.9756	1.4941
7	**9.0071**	**3.1882**	**0.4121**	**0.0893**	**0.2791**
8	15.4071	13.6979	1.7304	1.9612	2.1732

Best results overall are achieved by experiment #7, which is beaten by experiment #5 in some metrics. Both configurations show similarly good results.

A summary of the details of each experiment can be checked in the following table. All experiments share optimizer (Adam), loss function (MSE), metric (MAE), batch size (32), epochs (50, with early stopping) and patience (5)

Experiment	Backbone	Layers	Training steps	Learning rates
1	ResNet50	Classification	1	1e-5
2	Inception ResNetV2	Classification	1	1e-5

3	ResNet50	Classification Dropout FCC FCC	1	1e-5
4	Inception ResNetV2	Classification Dropout FCC FCC	1	1e-5
5	ResNet50	Classification Dropout FCC FCC	2	1e-4 1e-5
6	Inception ResNetV2	Classification Dropout FCC FCC	2	1e-4 1e-5
7	**ResNet50**	**Classification** **Dropout** **FCC** **FCC**	**3**	**1e-4** **1e-5** **1e-5**
8	Inception ResNetV2	Classification Dropout FCC FCC	3	1e-4 1e-5 1e-5

Some experiments playing with the value of some of the unrefined hyperparameters were performed but are not mentioned in this report because of the lack of conclusions drawn from them.

3. FINAL REMARKS

In this work we have tested different backbones, leading to ResNet50 providing the best results. At first, adding fully connected layers helped reduce the MAE metric but did not have a great impact on bias mitigation. In that sense, adding training steps worked the best, as if the whole structure focused more on specializing itself in each one of the classes. We used class weights to reduce the imbalance in representation of different groups in the dataset.

For future work, more backbones could be explored. Also, we should dig deeper into the search and tuning of hyperparameters such as batch sizes, learning rates, etc.