OPT 2 – COMPUTER VISION REPORT
for Optional Task

October 23, 2022

Group members:

Salma Ouardi, salma.ouardi@universite-paris-saclay.fr, sallyxhope,
Ni Wang, niwang.fr@gmail.com, Ni_Wang

1 Summary of contributions

In the first task of this project, we approached the competition 'Age perception Challenge' through two strategies, data augmentation and custom loss. In comparison to the default backbone ResNet50, another model, namely 'NASNet' (Neural Architecture Search Network), was implemented as backbone for training. The results of both two strategies were analysed. The performance of NasNet was evaluated in the presence of two sets of hyperoarameters.

In the optional task of this project, we did some hyperparameter tuning to the model, changed some parameters like the optimizer, the batch size and such, and we were able to get some results that are better.

1.1 Data augmentation strategy

Two major categories of methods were implemented:
Approach 1. Fast and flexible image augmentation with library Albumentations.

The process of image data augmentation mainly consists of four steps: Import necessary modules and packages to read images from the disk (e.g., OpenCV), define an augmentation pipeline, read images from the source, pass images to the augmentation pipeline and receive augmented images.

To define an augmentation pipeline, we created an instance of the Compose class. As an argument to the Compose class, we passed a list of augmentations we want to apply. A call to Compose will return a transform function that will perform image augmentation. With this library [1], a few commonly used image transformation methods can be easily called.
Approach 2. Manually creating image transformation class for the purpose of augmentation.

Different image transformation methods are to combined to created augmented image datasets, including: horizontal/vertical flipping, contrast changing, saturation changing, grayscale, brightness changing, and so on.

1.2 Custom loss strategy

Custom loss is explored by adding weights to each label while calculating the MAE. We kept all other parameters to the notebook default i.e., we used a resnet50 model pretrained on face dataset, and trained it using a mean squared error loss function with a learning rate of $1e-5$.

To calculate this weight, we have experimented with several different averaging methods surrounding arithmetic mean, geometric mean, and harmonic mean. We find results with non-weighted and weighted means of the mentioned methods. For weighted means, we employed both hard-coded methods and dynamic weight calculating methods. For our hard-coded method, we based the weights on our observations that certain attributes such as age had the highest bias and thus, we assigned it a weight of significant weight compared to other attributes. Our other weighing method was dynamic in the way that we calculated the weight from the bias during each training iteration.

In all the experimented averaging methods, we found that harmonic mean gave us the best results. We tabulate our results below.

1.3 Training strategy

Here we consider NASNet-A [2], the highest performance model that was found for the CIFAR-10 dataset, and then extended to ImageNet 2012 dataset, obtaining state of the art performance on CIFAR-10 and ImageNet 2012.

A pretrained model that has been trained on a subset of the ImageNet database. which is a family of models that were designed automatically by learning the model architectures directly on the dataset of interest.

2 Experiments & Results

2.1 Data Augmentation

An example of augmented image dataset is displayed in Figure1 Through data augmentation, significant improvements

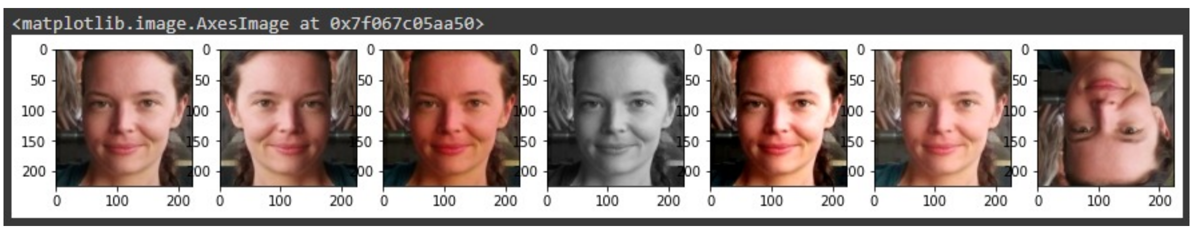

Figure 1: An augmented image dataset

of age estimation results have been found in both individual implementation and combination of different methods, displayed in Table2.1.

Transformations	MAE	Age Bias	Age Bias (Group 4)
Zoom and Shift	10.23	5.22	17.76
Flip, brightness, blur, translation and zoom	9.43	3.56	15.12
horizontal + vertical flip + brightness + ..	9.55	3.99	16.28

Table 1: Different transformation techniques and the outcome

2.2 Custom Loss

Results of the strategy of custom loss implemented on Validation dataset and Test dataset are displayed in Table2.2 and Table2.2 respectively.

2.3 Hyperparameters Tuning

The training history of the default model ResNet50 is displayed in Figure2.

NasNet	Trial 1	Trial 2
input˙shape	(224,224,3)	(224,224,3)
weights	imagenet	imagenet
include˙top	False	False
pooling operation	GlobalAveragePooling2D	GlobalAveragePooling2D
activation	sigmoid	sigmoid
optimizer	Adam	SGD
loss	mse	mse
metrics	mae	mae
batch ˙size	64	32
epochs	100	30
learning rate	1e-4	1e-4
callbacks	monitor='val˙loss', mode='min'	same
shuffle	True	True
verbose	1	1
Results	val_MAE = 0.0678; Val_Loss = 0.0080	Val_MAE = 0.09 val_loss = 0.01

Table 2: Changing Hyper-parameters in NasNet for better results

Validation Set	Age	Age + Gender	Age + Gender + Ethnicity
MAE	11.325	10.741	7.807
Age Bias	2.134	c	5.118
Gender Bias	0.018	0.018	0.059
Ethnicity Bias	1.481	0.659	0.883
Expression Bias	0.605	0.649	0.428

Table 3: Results on validation dataset

From the baseline of ResNet50, this section aims to figure out another powerful neural architecture in this task. Another type of model from the category of NestNet, 'Nasnetmobile', is applied as a backbone with the help of image-classifiers package. The newly chosen backbone was evaluated on two sets of different hypterparameters. The results from those two sets of parameters (namely 'trial1' and 'trial2') were compared in Table2.1.

3 Final Remarks

3.1 Experimentations remarks

- With the second approach of data augmentation (manually creating image transformation class), the MAE was found to be 4.2.
- In the strategy of data augmentation, it was discovered that a combination of various data augmentation methods delivered better results than using each of them individually.
- During the process of tuning hyperparameters, it has been discovered that NasNet can serve as a powerful backbone for age estimation task given good combination of hyperparameters. Two optimizers have been tested together with various learning rates. SGD optimizer appeared to be slightly hard to tune as the learning curve was either very steep or flat. Adam optimizer was selected as the best one and used with a bigger learning rate than in the initial training. At the same time, it has been found that NasNet is sensitive to batch size.

Test Set	Arithmetic	Geometric	Harmonic	Weighted Arithmetic Mean	Weighted Harmonic Mean
MAE	7.943	8.735	7.26	8.778	8.377
Age Bias	6.312	8.745	3.41	6.830	5.084
Gender Bias	0.333	0.288	0.571	0.224	0.012
Ethnicity Bias	1.088	1.228	0.838	1.130	1.202
Expression Bias	0.749	0.527	0.460	0.211	1.187

Table 4: Results on test dataset

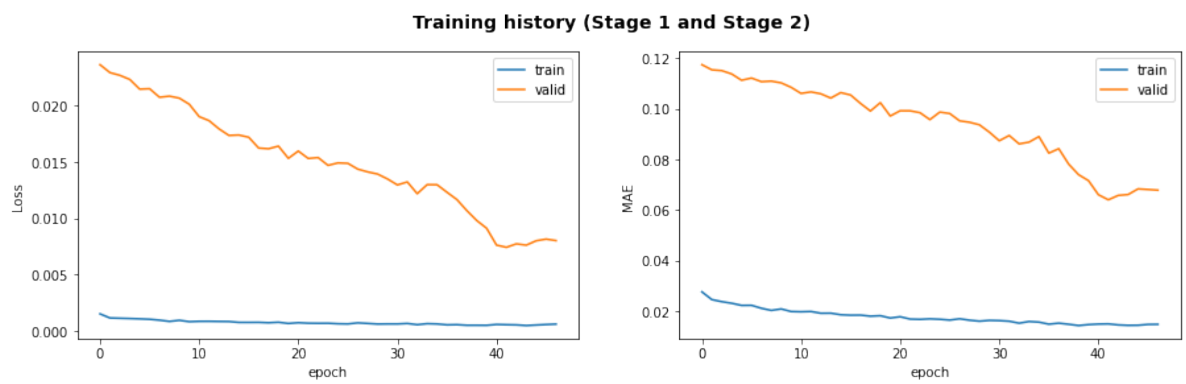

Figure 2: Training history of the model

- Through strategy of custom loss, not much improvement expected since genders are quite balanced, while MAE = 10.74.

3.2 Limitations

- One of the shortcomings of this project so far is a lack of comprehensive comparison among sufficient sets of hyperparameters. For example, in the evaluation of NasNet, only two sets of hyperparameters were compared. A reason of this lack is that Google Colab does not allow infinite times of running of program to test out.
- Another shortcoming of this project is that we only report results using MAE as loss metric. However, we tried Huber loss as well but found little improvment in performance. Furthermore, we would have liked to use AUC since we have read literature that claim that it improves learning in an unbalanced dataset. However, we were restricted in our efforts by time and compute resources.
- At the same time, no image re-scaling has been attempted before feeding datasets into the architecture. The lack of this step has resulted in no acceleration of training procedures.

References

[1] Alexander Buslaev, Vladimir I. Iglovikov, Eugene Khvedchenya, Alex Parinov, Mikhail Druzhinin, and Alexandr A. Kalinin. Albumentations: Fast and flexible image augmentations. *Information*, 11(2), 2020.

[2] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V. Le. Learning transferable architectures for scalable image recognition, 2018.