UNIVERSITÀ DEGLI STUDI DI SALERNO

Dipartimento di Ingegneria dell’Informazione ed Elettrica e Matematica applicata

Corso di Laurea in Ingegneria Informatica

Artificial Vision: Final Project

Gruppo 24:	Matricole:
Leonardo Galiano	0622701608
Antonio Iannaccone	0622701557
Nicolò Grieco	0622701578
Carmine Fruncillo	0622701615

ANNO ACCADEMICO 2021/2022

- DESCRIPTION OF THE PROBLEM:

The goal of this project is to realize a method to achieve the simultaneous classification of three facial attributes: beard, moustache, glasses. The three classification problems are binary, since the neural network is asked to determine the absence 0 or the presence 1 of the specific facial attribute. In particular, the problem is a Binary Multi-label classification, since each face is labelled with three binary labels: beard: (0 absence, 1 presence), moustache: (0 absence, 1 presence), glasses: (0 absence, 1 presence).

- ARCHITECTURE OF THE CLASSIFIER:

For the solution of the project we have choosen a deep neural network(VGG Face) based on ResNet50. The network presents a final layer with three output neurons used for the classification of the three classes Beard, Moustache, Glasses.

➢ Design Choices:

Instead of using traditonal method (doing feature extraction and then feed this to a SVM), we use a deep learning network that solved this problem learning automatically features. At the beginning we think that use the traditional method was the best way because the utkface dataset giving to us is composed of 9770 samples, not a lot. Furthermore the dataset was also imbalanced and so we decided to do data augmentation obtaining amount 16000 images. Also if this number of samples is very small for deep networks this satisfy the classification of beard, moustache and glasses with acceptable results.

- DATA PRE-PROCESSING E DATA AUGMENTATION:

Labels(Beard,Moustache,Glasses)	Quantity
0,0,0	8086
0,0,1	864
0,1,0	184
0,1,1	67
1,0,0	419
1,0,1	150

Here is represented the division of the dataset for each labels. The total samples of the dataset is 9770. We can note that the dataset is inbalanced because, for example, for the class 0,0,0 there are more samples of the other classes. So we have used some techniques of data-augmentation to increase the numbers of samples and to balance the dataset.

- Tecniques of Data-Augmentation:

  As written before the numbers of samples for the class 0,0,0 are more of the other classes so we have applied data-augmentation only on the remaining classes. The tecniques are this:

  - Brightness control
  - Blurring addition
  - Rotation of 30° in different directions

  Furthermore we have used other data-augmentation techniques but only on specific classes. For the class:

- 1,0,0 (beard only): rotation of 60°

- 0,1,0(moustache only) and 1,0,1(beard and glasses): rotation of 60° and 90°

- 0,1,1(moustache and glasses): rotation of 60°,90° and 180°

- Final Dataset:

  The final dataset after the operations of data-augmentation above described is composed of 16122 samples.
  For the training procedure we have shuffled and splitted the dataset in 80% for the training set and 20% for the validation set.
  So we have 12897 images for the training set and 3225 for the validation set.

TRAINING PROCEDURE:

- The design choices for the training procedure are based on the following aspects:

- The dataset has been splitted into 80% for the training set and 20% for the validation set.

- The used optimizer is Adam. Adam optimization is a stochastic gradient descent method. This is a good method for problems with a lot of data.

- The loss function is the binary cross-entropy. This function is perfect to solve binary problems (0 absence, 1 presence).

- The learning rate used is 1e-4.

- The use of batches: for each epochs does not use the entire dataset on the training but only a small part.

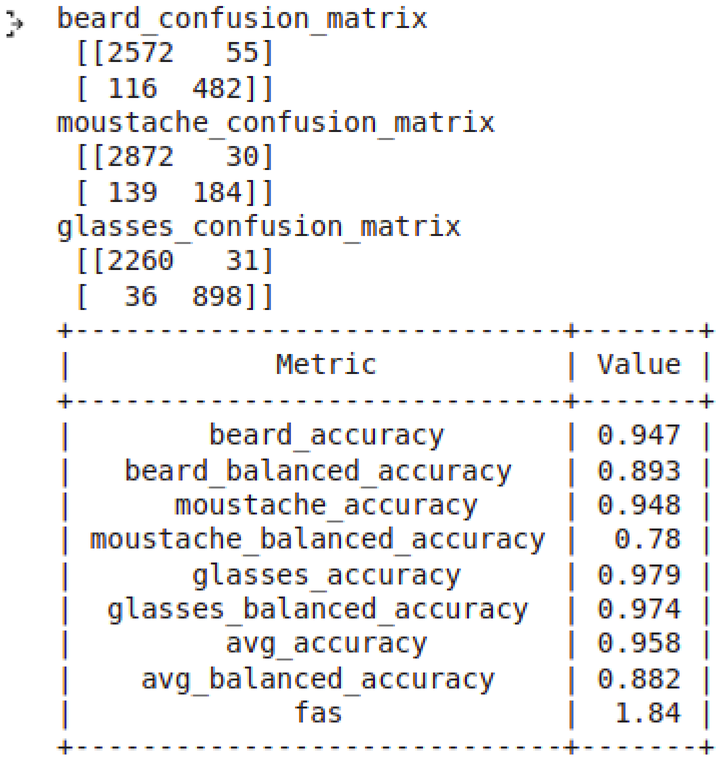
- Early stopping is a form of regularization used to avoid overfitting during the training. We have used a patience of 10 epochs.

- QUANTITATIVE AND QUALITATIVE ANALYSIS

  At the beginning we thought that use the traditional method was the best solution. It consist in doing feature extraction with LPG and HOG, then use PCA, and then use a traditional machine learning algorithm. But then we realized that this method can be replaced with the use of a deep neural network.
  In particular we have tested two network based on Resnet50: one with ImageNet weights and another one which uses VGG-Face weights. We have tested these network on the validation set used as test set.

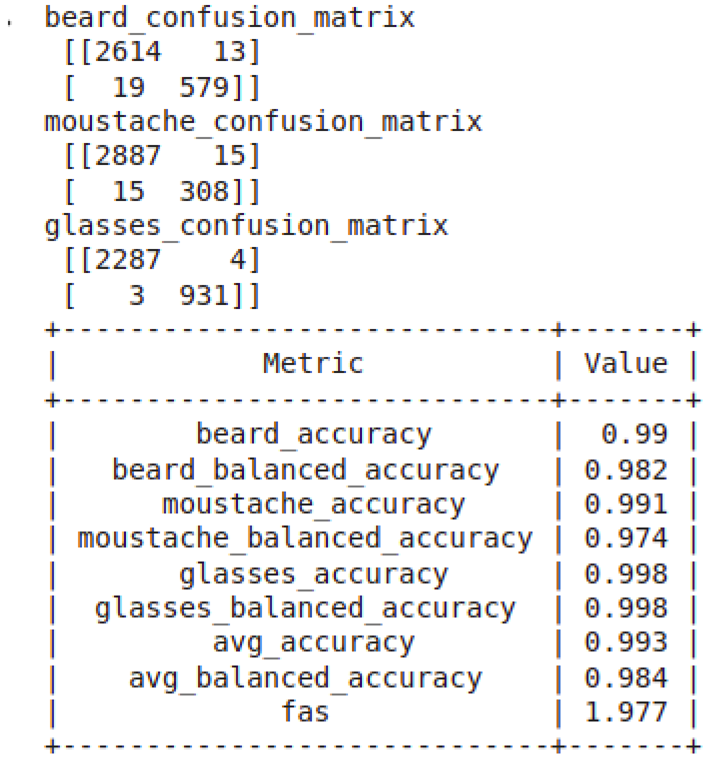
- Quantitative Analysis:

  - These are the results of the deep neural network based on Resnet50 with ImageNet weights.



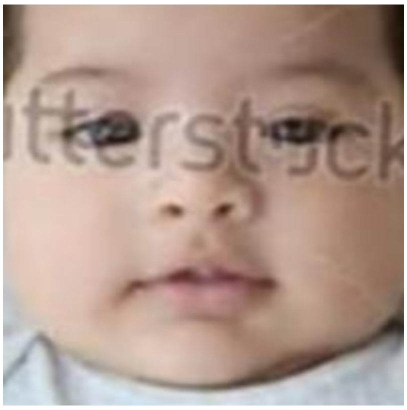
```
beard_confusion_matrix
 [[2572   55]
 [ 116  482]]
moustache_confusion_matrix
 [[2872   30]
 [ 139  184]]
glasses_confusion_matrix
 [[2260   31]
 [  36  898]]
+------------------------------+-------+
|            Metric            | Value |
+------------------------------+-------+
|        beard_accuracy        | 0.947 |
|    beard_balanced_accuracy   | 0.893 |
|      moustache_accuracy      | 0.948 |
|  moustache_balanced_accuracy |  0.78 |
|       glasses_accuracy       | 0.979 |
|   glasses_balanced_accuracy  | 0.974 |
|         avg_accuracy         | 0.958 |
|     avg_balanced_accuracy    | 0.882 |
|             fas              |  1.84 |
+------------------------------+-------+
```

  - These are the results of the deep neural network based on Resnet50 with VGG-Face weights.

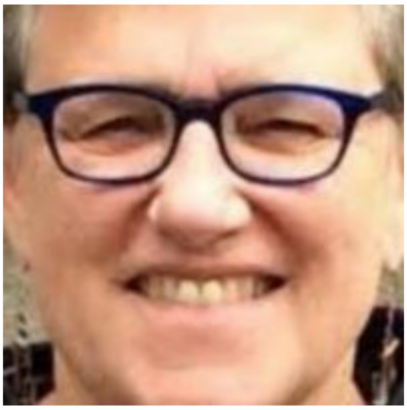


```
beard_confusion_matrix
 [[2614   13]
 [  19  579]]
moustache_confusion_matrix
 [[2887   15]
 [  15  308]]
glasses_confusion_matrix
 [[2287    4]
 [   3  931]]
+------------------------------+-------+
|            Metric            | Value |
+------------------------------+-------+
|        beard_accuracy        |  0.99 |
|    beard_balanced_accuracy   | 0.982 |
|      moustache_accuracy      | 0.991 |
|  moustache_balanced_accuracy | 0.974 |
|       glasses_accuracy       | 0.998 |
|   glasses_balanced_accuracy  | 0.998 |
|         avg_accuracy         | 0.993 |
|     avg_balanced_accuracy    | 0.984 |
|             fas              | 1.977 |
+------------------------------+-------+
```

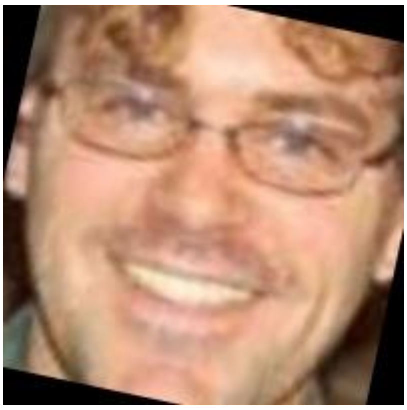
- Qualitative Analysis:



This image have groundtruth [0,0,0] but the neural network predict [0,0,1]. This image have a poor resolutions so the classifier makes a wrong prediction.



This image have groundtruth [0,0,1] but the neural network's prediction is [0,0,0]. Sometimes it's possible that classifier makes wrong prediction on particular type of glasses.



This image have groundtruth [0,1,1] and the neural network's prediction is [0,1,1]. In this case our method is right referring to the groundtruth but this one is wrong.

- Final Considerations:

  The single multitask classifier with 3 outputs is our solution. The tree outputs are beard, moustache and glasses. With this implementation we have obtained optimal results. This is possible also because the network we have choosen is specialized in face analysis. Furtheremore the pre-trained weights on VGG-Face concure.