Lab Assignment

David Barbas Rebollo

Computer Vision
June, 2021

1 Introduction

The objective of this assignment is to solve different computer vision problems achieving certain goals for each individual task. These tasks make reference to the different concepts taught in class.

In each task the goal as well as the solution given is explained. The problems tackled on this project are Gender Recognition, Car Model Identification with Bi-linear Models and Image Colorization.

2 Gender Recognition (3 points)

As the title of the task states the objective of this task is to classify different images by gender. The dataset used consists of 100x100 rgb images from the "Labeled Faces in the Wild" dataset in realistic scenarios, poses and gestures. There are 10,585 available for training, while only 2648 for test.

The goals of this task are to do a first implementation with >97% accuracy over the test set and a second implementation with >92% accuracy over the test set but with a model with less than 100K parameters.

2.1 Implementation

Firstly, to achieve a more robust training several data augmentation parameters are considered such as image shifting, horizontal flip, zooming, image rotation and shearing. Secondly, for a smoother training over 200 epochs a learning rate scheduler was used for the SGD optimizer.

Secondly, the topologies considered were smaller versions of the VGG, Resnet and Densenet topologies. Having 7.8M, 7.8M, 420K parameters respectively. Therefore for experimentation smaller versions of the previously proposed topologies were considered. Having 94K and 99K parameters respectively.

The results obtained of the experiments are shown below:

Table 1: Results of experimentation for the Gender Recognition task.

Model	# of Parameters	Accuracy (%)
Mini VGG	94K	97.89
Mini Resnet	99K	97.81

From the results obtained we can conclude that the complexity of network for this task does not need to be very high. Also, residual connections do not have an important role on this task. However, the most important feature is a good data augmentation method to help training. Which is why the goals of these task could have been achieved.

3 Car Model Identification with Bi-linear Models (5 points)

The second problem involves classifying images 20 different car models via a bi-linear network. These are 250x250 rgb images. They are divided into 791 images for training and 784 for test.

The goal of this task is to obtain >65% of accuracy over the test set, learn how to name the layers of a network and connect models with the outer product operation.

3.1 Implementation

Firstly, to achieve a more robust training several the same data augmentation procedure as for the previous task is used. However, for this the learning rate scheduler used was smoother and the Adam optimizer was used instead.

There were 2 different topologies used for this task, a custom small VGG and a pre-trained VGG16. For the custom VGG was trained for 200 epochs with learning rate annealing and then trained again for 200 epochs with a learning rate of 0.0001. The pre-trained VGG16 was trained with all VGG16's layers freezed for 200 epochs with learning rate annealing. Then these layers were unfreezed and trained for 200 more epoch with a learning rate of 0.0001.

The results obtained of the experiments are shown below:

Table 2: Results of experimentation for the Gender Recognition task.

Model	Accuracy (%)
Custom VGG	7.65
VGG16	75.63

From the results obtained we can conclude that the learnt weights from the imagenet task favor considerably the recognition ability of the network. As there are few available training images the network is able to finely recognize the different car models better than expected. Finally, the results obtained achieve the objectives proposed by this task.

4 Image Colorization (3 points)

The final problem of this lab assignment consists in developing a network capable of learning how to colorized black and white images. The goal is to understand the alpha implementation and ultimately, use an encoder-decoder network merged with the InceptionResNetV2 model. Also, understand its implementation and how all image operations are performed for a correct training.

It is important to notice that to enable the usage of all samples without loading all simultaneously in memory *keras' flow_from_directory* method was used. Other methods were as well adapted to these changes.

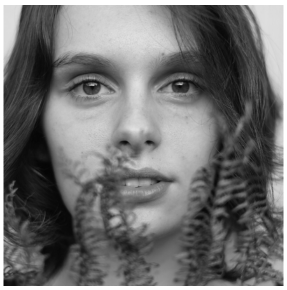
4.1 Alpha Version

The objective of this version is to familiarize the user with the syntax of the model and enabling to see a result in a few minutes.

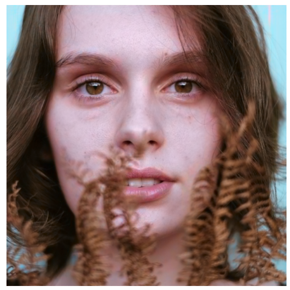
It also introduces how images are transformed from RGB to Lab. L stands for lightness, a and b for the color spectrum of green-red and blue-yellow. Therefore, given black and white images are only represented with lightness.

The model only has to learn to predict 2 channels instead of 3. All these concepts are used in the posterior version.

The alpha version only learns to predict one image. Its results can be seen below:



(a) Black & white image.



(b) Colorized image.

(c) Original image.

4.2 Full Version

The final version enables to generalize colorization and has four components. The encoder, decoder, the classification from InceptionResNetV2 and the fusion layer. The fusion layer merges the encoded image with the classification of the pretrained classifier. The image below shows how the four elements merge together:

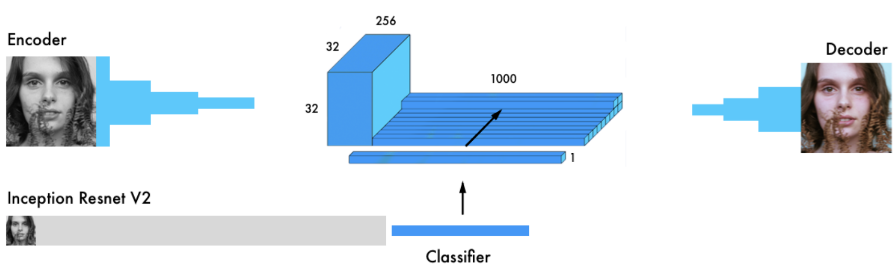


Figure 2: Network's architecture.

By transferring the learning from the classifier to the coloring network, the network can get a sense of what is in the picture. Therefore, enabling the network to match an object class with a coloring scheme.

The final version was trained over 50 epochs 9K training images, 500 for validation and another 500 for test.

The results obtained over the test set are not able to resemble the realistic photographs with the actual training. However, revising the results the network was generally able to identify green and blue backgrounds such as green scenery and the sky. Also the output images normally have a pale brown tonality resembling a sepia or an instagram filter. Therefore people with darker skin color have a more realistic colorization.

Some examples of the best results obtained by this network are: