Light CNN applied to CASIA-FaceV5-160

Giorgia Nidia Carranza Tejada i6231006

May 2020

1 Introduction

One of the tasks covered by computer vision is face recognition. To be able to identify distinct faces, the process can be divided into several phases such as the detection, the alignment, the representation, and finally the classification. This project will focus on the representation part, which aims to extract the features points from the faces.

The motivation to use the approach Light CNNN proposed by Xian[3] is related to the faster and small dimension of CNN, which is against the state of the art that tends to increase models reducing its portability. Therefore, the experiments used a single core. Despite the size of CNN, it presents good results in terms of accuracy, precision compared to other methodologies.

2 Related Work

Regarding the extraction of features for face verification by using CNN, a pioneer DeepFace[2] performs a 97.35 percentage of accuracy over LFW by combining several networks. Instead, FaceNet[1] archives a performance of 98.87 with no alignment and 99.63 with it. Finally, the model used Light CNN-29 reached a 99.33 percentage of accuracy.

3 Methodology

The experiments were executed through the Light CNN-29 model proposed by the paper[3]. It presents considerable results in comparison to state of the art concerning the LFW data set. The implementation of the model is contained in AlfredXiangWu's GitHub project [1]. Some modifications were realized to the code to run the experiments, such as the number of classes differs from their training test.

The paper explained that the model was trained and realized most of its experiment on the common dataset used for face recognition, such as CASIA-WebFace, MS-Celeb-1M, LFW, MegaFace, IJB-A, etc. These datasets available

[1] `https://github.com/AlfredXiangWu/LightCNN.git`

large-scale face datasets are mainly Westerners, containing few Asians personages. My intention was the employment of a dataset containing images of people of Asian ethnicity, to verify if the somatic traits may influence the performance. The datasets available only with these constraints are limited since some are not publicly available. Others are labeled for a different purpose, such as the recognition of the age, instead of the person. Besides, deep learning schemes required a large amount of labeled data, but the majority datasets accessible had few examples in comparison to the previously mentioned.

In this case, the dataset used is CASIA-FaceV5, which contains 2,500 color facial images of 500 subjects retrieved by another Github project[2]. As can be noticed, the dimensions of the examples were notably reduced from 10k to only 500, so this component should be taken into account during the analysis of the performance.

Therefore, some experiments were executed to deal with this problem. The first consisted of the usage of the pre-trained model given by the GitHub repository of the LightCNN-29Layers depending on the idea contained in the fourth section of the paper. They did not re-train or fine-tune the Light CNN model on the other testing database but, they directly extracted the features, and compute the cosine similarity. This method did not provide any result since it raised a runtime error caused by the different number of classes needed.

The second approach consisted of training the model with CASIA-FaceV5. It is further divided by, at first, train the model without any preprocessing techniques, then the second, by using it.

The preprocessing consists of reducing the size of the image, starting with the face recognition, and then increase the margin size to have approximately 128. Then, it was converted to grayscale, resize another time to 128 through cubic interpolation and normalized.

4 Results

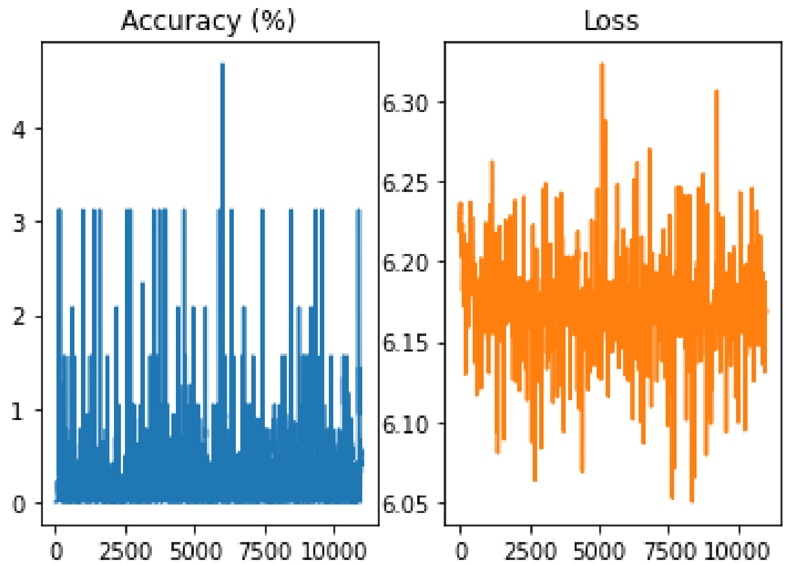
The experiment realized are four, divided into two with the preprocessing, and the other two with it. Each pair is composed of training of 200 and 500 epochs since values of accuracy and loss were considered far from a reliable performance.

The dataset was divided into train, validation, and test set by the following proportion 70%, 20%, and 10%.

4.1 Experiments without preprocessing

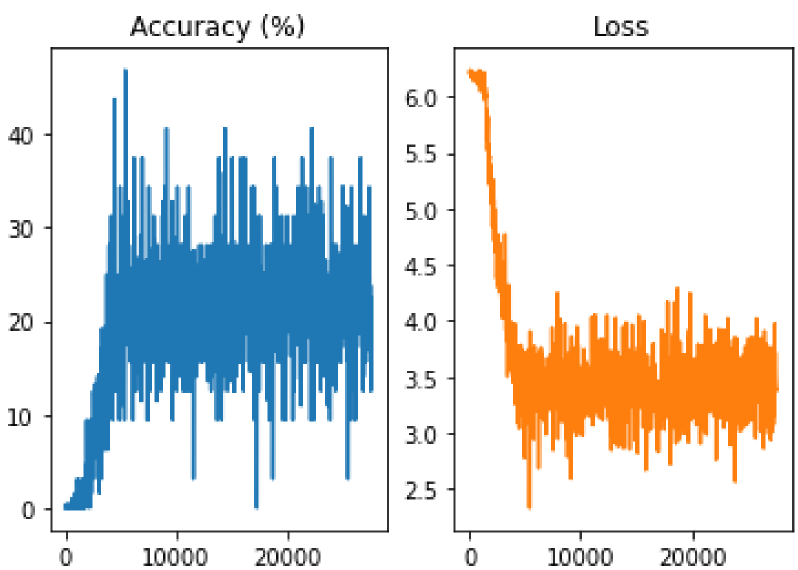
The first experiment consisting of 200 epochs obtains the following performance for loss and accuracy.

[2] `https://github.com/Mendru/facenet.git`



The result presents a poor performance with an average of accuracy at 2% and a loss stables to 6.2.

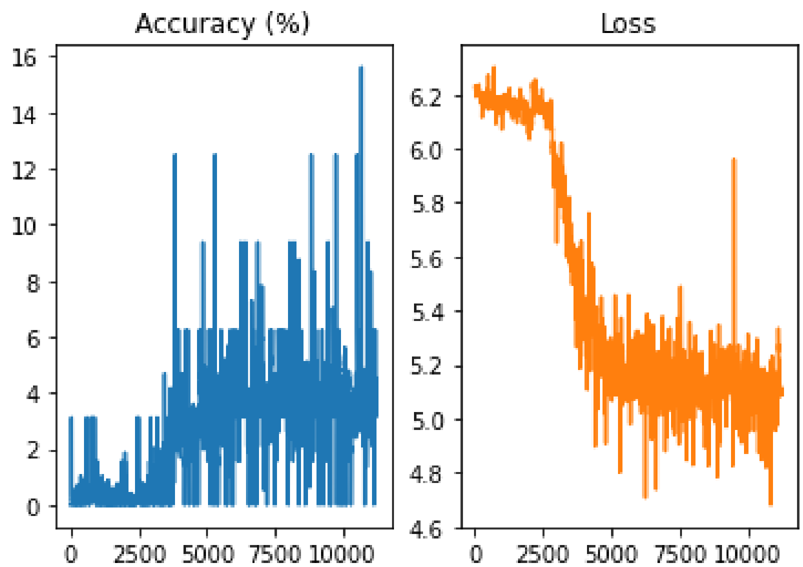
The second experiment trained the same model on 500 epochs. The outcomes of accuracy and loss are followed presented.



The average of losses remains under 4 instead, the accuracy varies between 20-30%.

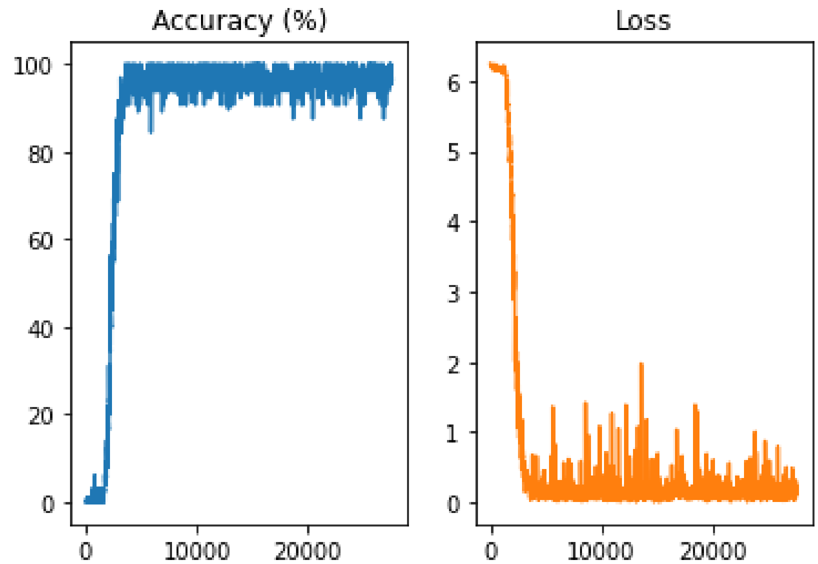
4.2 Experiments with preprocessing

The first experiment consisting of 200 epochs obtains the following performance for loss and accuracy.



Compared to the first experiment of subsection 4.2, it improves as the rates of the loss initially decrease to then be stable to 5.3. Also, the accuracy, previously stable under 3%, increases to an average of 6%. However, the levels of both marks are extremely low, even with the preprocessing.

To analyze if some improvement can be achieved by increasing the training iterations, the second experiment consisting of 500 epochs. It presents the following results.



In comparison to the second experiment of subsection 4.2, the employment of the preprocessing considerably augments the rates of the loss, which remains stables under 1. Also, the accuracy, previously to 30%, increases and persist constantly between 90-100%. However, the results of the last validation did not confirm the remarkable performance since it results in an average accuracy of 25% and an average loss of 7.789.

5 Discussion

The model reveals was trained using a GPU provided by Google Colab. The training of the dataset CASIA-FaceV5-160 was relatively fast, implying 2/3 hours but, it can be related to the small dimension of examples. Instead, to train the model with CASIA-WebFace, the duration increases to more than 12 hours for less than 50 epochs. The results were not exposed in this paper since the training was set to 80 epochs, therefore the information was deleted with the disconnection of the Google Colab.

Besides, the aim of the project was the employment of deep face representation over a different dataset from the commonly used for train and test. To prove if the model can be considered biased, the data applied derived from CASIA-FaceV5-160, consisting of images of people of Asian ethnicity. The research of an appropriate dataset for this purpose was difficult since many of the proposal databases contain, as described, a majority of Caucasian people. Also, the dimension of the suitable data was considerably small than the size needed for deep learning models. Even CASIA-FaceV5-160 has a size of only 501 categories in comparison to 10K of CASIA-WebFace.

This issue reflects on the performance of the model since the results obtained were extremely low in terms of loss and accuracy during the validation test, even if the data were preprocessed before being used. To be compared to its state of the art in this topic, the model should be trained by a dataset of at least 80K, as the original.

6 Conclusion

The project aspires to establish if the model contains bias in its features extraction caused by the insufficient diversity of the original dataset. However, it was not possible to answer the question due to the lack of large-scale data set devoted to another ethnicity not Caucasian. A possible future work could imply the application of data augmentation to increase the size of the collection.

References

[1] Florian Schroff, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering". In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2015, pp. 815–823.

[2] Yaniv Taigman et al. "Deepface: Closing the gap to human-level performance in face verification". In: *Proceedings of the IEEE conference on computer vision and pattern recognition.* 2014, pp. 1701–1708.

[3] Xiang Wu et al. "A light cnn for deep face representation with noisy labels". In: *IEEE Transactions on Information Forensics and Security* 13.11 (2018), pp. 2884–2896.