AI Report & Deep learning: TP7-

Camille FAHY
AI and Deep ESIEE - 2024
France
camille.fahy@edu.esiee.fr

Hugo KOTHE
AI and Deep ESIEE - 2023
France
hugo.kothe@edu.esiee.fr

Abstract- Facial recognition is one of the most studied research specialities in the field of computer vision. It is used to authenticate or identify a person based on their face. This technology can be applied to many fields, but it faces challenges such as the amount of data needed to obtain a high-performance model, the complexity of the computation, overfitting, etc.

Index Terms-Deep learning, face recognition, face detection, face encoding, accuracy, overfitting

I. Introduction

Facial recognition is one of the most widely studied research specialities in the field of computer vision. It is widely used to identify individuals, whether for security reasons, on social networks or in medicine.

The first stage in facial recognition is face detection. Each face is then framed. However, it must be possible to detect a face even if it is badly oriented, under different positions, luminosities or scales. It is therefore necessary to align the faces. In addition, the computer must take the marks that are most relevant to it in order to measure the faces: this is known as encoding. Finally, all we have to do is match the measurements of test image to those of the images in our database.

Using this method, we were able to build and train a high-performance model for identifying people using their facial image, and to train it on our own set images. However, this required importing a large amount of data, a model with significant computational complexity and the adjustment of hyperparameters.

II. Related Works.

One emerging area of facial recognition is disease diagnosis. These technologies could improve the detection of certain diseases. One example, cited in the paper, is endocrine diseases, which are characterised by typical changes in the facial features of bones, muscles and soft tissues. "The general symptoms of these diseases at an early stage are easily confused with other metabolic syndromes. The reference diagnostic process is complex, with several cases of hormonal tests and imaging examinations [26 , 27]. The high diagnostic performance of facial recognition has made it a rapid and accessible screening approach. Due to the increased release of growth hormones, people with acromegaly often have a rectangular face, enlarged nose and lips, prognathism and a rounded forehead [**26** . Various algorithms have been developed to detect the faces of patients with acromegaly. In 2006, Learned-Miller et al [**28** proposed a 3D morphable model to classify the frontal face into different categories semi-automatically. Forty-nine

patients acromegaly and controls were identified with an accuracy rate of 85.7%. "[1]

Facial recognition is also used in IoT.[2] This study presents a facial recognition system for controlling the opening of a door using a raspberry pie. This project uses deep learning. 5 images of 5 different people are taken by a camera to create the training set. The images are then augmented to give a total of 2,500 images, then cropped and reduced. It uses the 8-layer AlexNet architecture. The training initially comprises 100 epochs and is repeated with 20 epochs after the test phase. performance of this model is good, but it takes a long time to process the image. This project appears to be similar to our project and provides an example of a security application.

Facial recognition can be used for emotion detection. [3]

In this study, pre-processing consists of cropping the eyes and mouth using the Haar Cascades method. The training database contains 20 images of 6 emotion classes. To test the images, this article adopts the JAFFE database. The database contains 213 images (each image: 256*256 pixels) of Japanese women's faces, each image being defined in the original expression. There are 10 people in the expression library. The performance of this model on the test images gives results between 85 and 95% depending on the different emotions.

.

III. PROPOSAL

Here you will need to write about each of the topics that you worked on TP7.

A. Face detection

Write a short first paragraph on what is face detection (try to cite 1 work that you got this information from) Write a paragraph discussing the topics of:

- Talk about the data you received for face detection (what are the dimensions, how many images are there, etc.);
- Why do you need to do face detection for this data for your work?
- What were the data pre-processing techniques you used ?
- How did you divide training and test ?
- Discuss about the small convnet you use, give a brief explanation of why you created it this way;
- Discuss about your small convnet performance on the test set;
- Does the result improve when you change : the number of layers, number of hidden units, learning rate initial value ?
- If possible include images (face before detection - after detection, performance graphs of your model, comparison between models, etc)

Fig. 1. Image caption.

Face detection is an essential step in all facial recognition algorithms. It includes face alignment, face recognition face analysis. The purpose of face detection is to determine the presence of faces in the image and to return the location of the image. Although this seems to be an easy task for humans, it can be difficult for computers, particularly because of face positioning, scale, facial expression or brightness. [4]

To train our facial recognition model we have 218 dimension 2 images of varying sizes of 6 different celebrities as input data.

First, we listed all the paths to the images in our set. Then we read set of corresponding images, convert the OpenCV image from BGR (blue, green, red) colour space to RGB (red, green, blue) colour space, extract only the face from the photo using the extract_face(image, model="hog") function. The 'hog_detector' tool allows us to detect faces and comes from the dlib library.
We noted that for some images no face was detected. This mainly concerned people wearing glasses or with their faces cut off.

We then stored these extracted faces and the person's name in the "dataset" list and in a file called "faces_dataset.pkl". Finally, we normalised the images (by dividing the pixels by 255), encoded the labels, and then stored them in a 'faces_dataset.pkl' file.

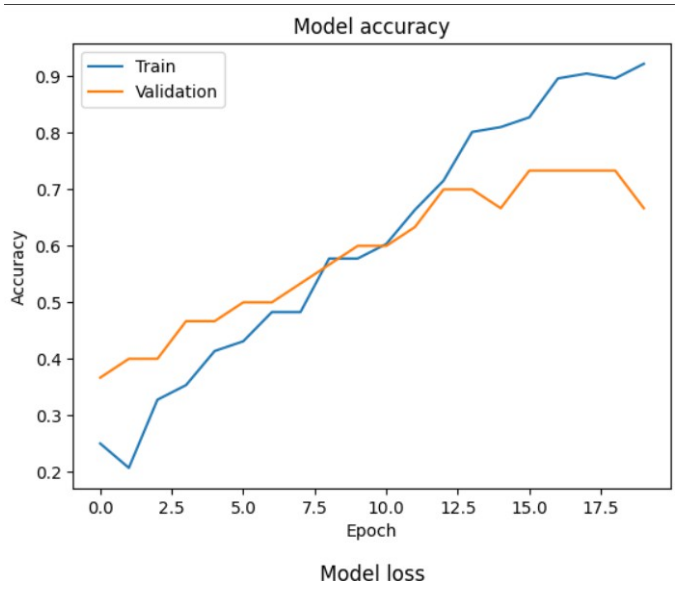
split the dataset into a training set and a test set using the function train_test_split(normalized_faces, integer_labels, test_size=0.3, random_state=42). The data is therefore divided into two parts: 70% training and 30% for testing.

We have built our model in several layers:

- a 2D convolution layer with 32 filters of size (3,3), and a ReLU activation function
- a max pooling layer with a window of size (2,2)
- a 2D convolution layer with 64 filters of size (3,3), and a ReLU activation function
- a max pooling layer with a window of size (2,2)
- a 2D convolution layer with 128 filters of size (3,3), and a ReLU activation function
- a max pooling layer with a window of size (2,2)
- a flattening
- a fully connected layer with ReLU as the activation function
- an output layer with softmax as the activation function for classification, in this case into 6 classes.

We have created a small convolution network which often used for classification tasks with a small dataset.

To improve the performance of our model and reduce overfitting, we then decided to add two Dropout layers with a rate of 0.2 after max Pooling. However, the performance of our model remains limited despite the adjustment of the hyperparameters. We obtained an accuracy of 70%.



.

B. Installation estimate

Write a short first paragraph what is Pose estimation (try to cite 1 work that you got this information from) Write paragraphs discussing the topics of:

- Talk about the data you are now using for pose estimation (what are the dimensions, etc.);
- Why do you need to do pose estimation for pose correction for this data for your work?

Fig. 2 Image caption.

modelPerformance.jpg

- What were the techniques you used for pose estimation ?
- How did you divide training and test ?

modelCompare.jpg

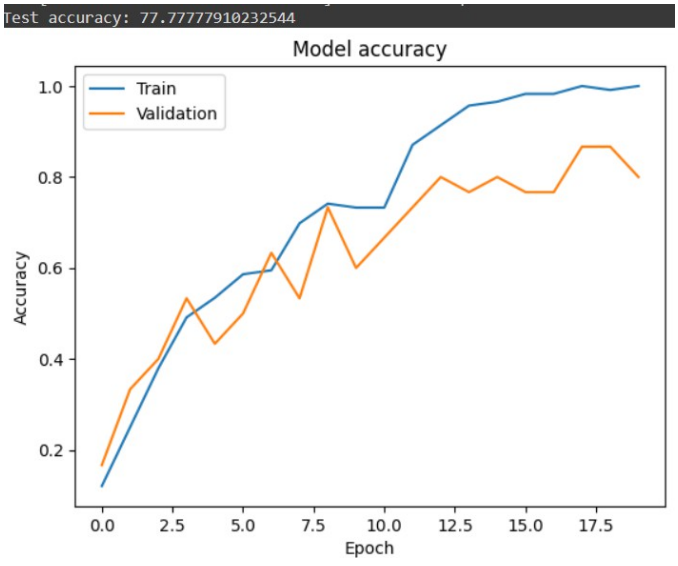
- Discuss about how your previous small convnet models performs on this test set;
- Does the result improve?
- If possible include images (New image sample of training image after pose correction- previous and after-, performance graphs, comparison between models, comparison between the performance of the model with this data and just the previous step, etc);
- Why do you think the results improved ?

Variations in facial pose or illumination are among the main problems in facial recognition. This is why pose estimation tools are very useful. Several techniques exist: those based on landmarks (Zhao and Gao 2006), which we will use in this tutorial, subspace methods based appearance, which treat the whole face as a vector of features in certain feature subspaces (Gong, 1996), or a combination of the two (Grundig and Hellwich, 2004). This allows faces to be aligned so that they are recognisable by the computer regardless of their pose. [5]

So by transforming our previously extracted faces we should be able to achieve better performance.

We applied the landmarks() function, which calculates a list of characteristic landmarks for each face. Then we apply the align_faces function. This aligns and centres our faces.

As before, we divided our data with 30% for the test and 70% for the training. We obtained better results than in the first part with a accuracy of 77.8%.



C. Face encoding

Write as the first paragraph what is Encoding - if you can be specific try to explain in the context of face encoding- (try to cite 1 work that you got this information from) Write paragraphs discussing the topics of:

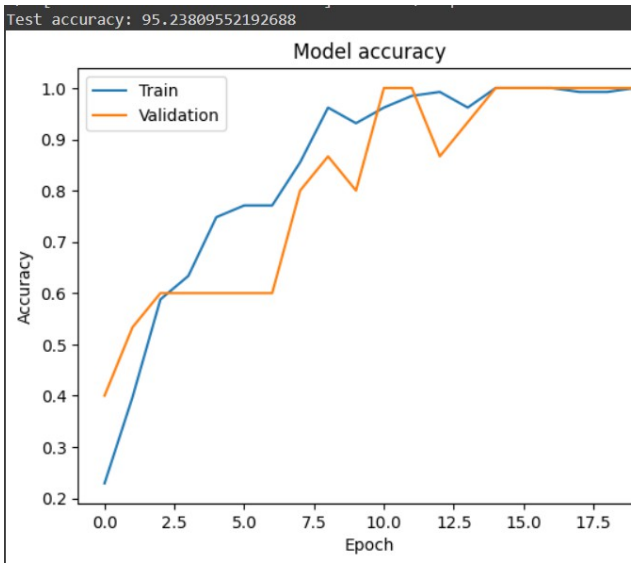
- What is the advantage of face encoding?
- What is the data that you are going to be using during face encoding?
- Discuss what the data becomes after face encoding; - Discuss about your training model for this new data;
- Does the result improve?
- If possible include images (performance graphs, comparison between models, comparison between the performance of the model with this data and just the previous step, etc);
- Why do you think the results improved (or why not)?

Encoding is a method used in facial recognition to determine the most important points to measure for each face, which together allow each face to be uniquely identified. The 128 points to be measured are determined deep learning methods. [6]

Unlike the models trained previously, we have size vectors (128) as input data. We are not focusing on all the details of the face, but only on the 128 measurements needed to identify it.

What's more, our model will be a neural network rather than a convolution network, because the image is no longer an image to be split into several parts, but a simple vector.

By encoding our images, the accuracy of our model increased to 95.2%, compared with 77% without encoding. Unlike before, we observed little overfitting. The validation curve follows the training curve. We could have increased the number of epochs to see how far the accuracy could be increased.



D. Face recognition

Write as the first paragraph what is face recognition (try to cite 1 work that you got this information from). What is the difference between face recognition and face detection ? Write paragraphs discussing the topics of:

- What are the classifiers you chose to do this task? Why ?
- Discuss about the performance of each classifier in this task;
- What is your best classifier ? Write about the performance;
- If possible include images (performance graphs, comparison between models, etc);

Facial recognition is a method of identifying faces based on their biometric characteristics. [7]

The difference between facial recognition and face detection is that facial recognition associates a face with a name, whereas face detection only detects the presence or absence of a face.

We tested 4 classifiers:

-logistic regression

-KNN

-SVM

-Neural Network

```
Logistic Regression: Accuracy = 0.9682539682539683, Training Time = 0.041774511337280
SVM: Accuracy = 0.9841269841269841, Training Time = 0.006528615951538086
kNN: Accuracy = 0.9841269841269841, Training Time = 0.016552209854125977
Neural Network: Accuracy = 0.9841269841269841, Training Time = 0.7190883159637451
```

Logistic regression is the least accurate. Despite a fast training time, the model is not the most suitable.

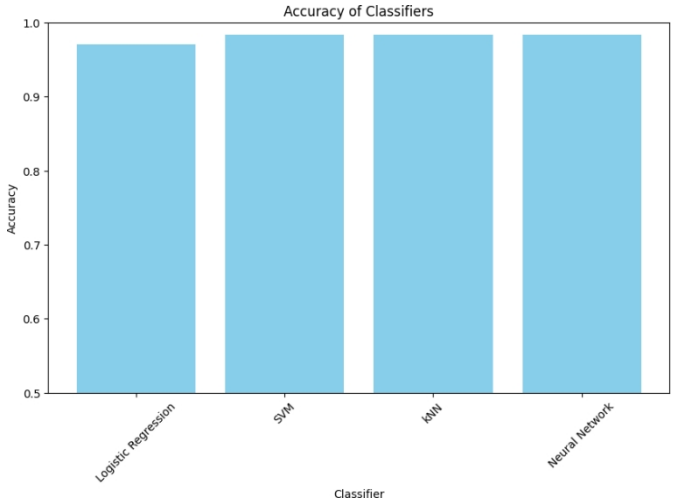
The Support-Vector Machine has very good accuracy (0.98) and a very short training time (0.0065), making it a very good classifier.

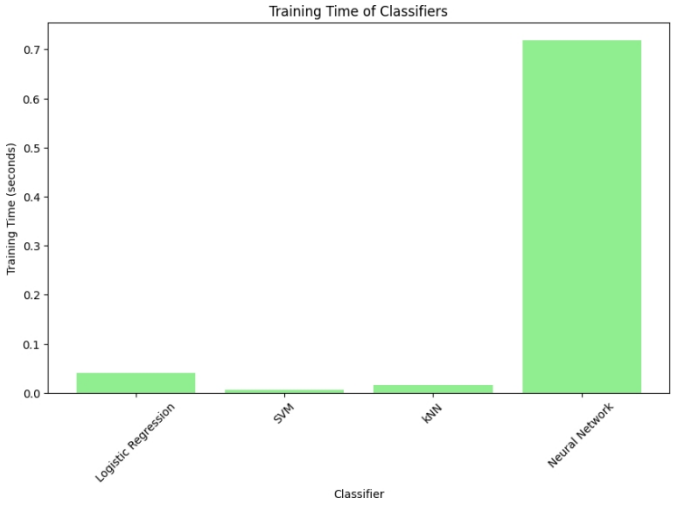
The k-nearest neighbour method also has good accuracy (0.98) and a short training time (0.016).

The neural network has good accuracy (0.98) but a long training time (0.7).

The most suitable classifier is the SVM (Support-Vector Machine).

We have noted that this is true when using Google colab, but that it is not necessarily true on all machines.





E. Personal dataset

- Talk about how you created your dataset; - Does your model work on your personal set?
- How is your model's performance in your dataset compared to the previous dataset? Is it good/bad ? (What is your opinion on why)
- Compare your dataset to the original data;

We took photos of ourselves and our families. We reduced the quality of the photos.

The performance of our model is quite good.

F. Extra - Bias analysis

- What is bias?
- Why is bias a problem in machine learning?
- Can you comment on a situation where bias can be a problem?
- Did you create a set trying to diversify your input examples ? (Is it all from the same person ?, Do you present different people ethnicities to verify if the model had bias towards some specific group (meaning better performed in one group)?
- Can you think of some statistics to calculate and present based on the original face detection data ? What about you data ?

[8] We might think that machine learning could be a solution for reducing inequalities and unfair decisions. However, machine learning decisions can also be biased. Indeed, when the machine learns by itself from examples, it is possible that certain data are biased, so the machine will incorporate these biases into the model. These biases can occur when the population in the training data is not representative of the overall population, or when the model is not complex enough to represent the data structure correctly.

Let's imagine that a company uses an AI algorithm to sort the CVs of job applicants. Here's how bias could be a problem in this context:

If the algorithm is trained on data that is based on prejudices such as age, origin or gender. The algorithm will be able to reproduce these biases in its decisions.

If the set of data used for training does not correctly represent all potential candidates, the algorithm may have difficulty deciding on profiles that are poorly represented.

If the selection criteria are based on stereotypes, this will be reflected in the decisions taken by the algorithm.

To try and reduce bias, we can analyse our training data and calculate the distributions according to several characteristics. For example, in our case, we could calculate the representation of different ethnic groups. We could also test our model on different subsets of people to ensure that performance is equal.

Biases therefore need to be taken into account as soon as the model is created.

CONCLUSION

Here you need to summarize in 1-2 paragraphs this work and what you learned.

In conclusion, this tutorial enabled us to learn more about face recognition and the associated deep learning techniques. We learned about the state of the art and the concepts of face detection, the problem of pose variation and encoding. We tried to improve the performance of our model using a number of techniques, such as drop-out, early stopping choosing the right classifier, etc. We applied all these methods to our dataset and obtained a high-performance model.

This practical course will be very useful for our future projects in artificial intelligence and will give us a better understanding of the current challenges in this field.

REFERENCES

[1] Review on facial recognition based application in diseases diagnosis' by Jiaqi Qiang, Danning Wu, Hanze Du, Huijuan Zhu ,Shi Chen , and Hui Pan

[2] 'IoT based facial recognition door access control home security system using raspberry pi' A. R. Syafeeza, M. K. Mohd Fitri Alif, Y. Nursyifaa Athirah, A. S. Jaafar, A. H. Norihan, M. S.
Saleha

[3] An Emotion Recognition Model Based on Facial Recognition in Virtual Learning Environment
Author links open overlay panel

D. Yang a, Abeer Alsadoon a, P.W.C. Prasad a, A.K. Singh b, A. Elchouemi c

[4] WIDER FACE: A Face Detection Benchmark' Shuo Yang1 Ping Luo2,1 Chen Change Loy1,2 Xiaoou Tang1,2

[5] HEAD POSE ESTIMATION IN FACE RECOGNITION ACROSS POSE SCENARIOS' M. Saquib Sarfraz and Olaf Hellwich,

[6] TP7 facial recognition Laurent NAJMAN

[7] W. Zhao, R. chellappa, P. J. Phillips, Face recognition: A literature survey, "ACM Computing Surveys (CSUR)",
December 2003

[8] Understanding Bias in Machine Learning Jindong Gu 1 , 2 , Daniela Oelke 2