DEMOCRATIC AND POPULAR REPUBLIC OF ALGERIA
MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH
UNIVERSITY El-BACHIR EL IBRAHIMI BORDJ BOU ARRERIDJ
FACULTY OF MATHEMATICS AND INFORMATIC
DEPARTEMENT OF INFORMATICS
MASTER IN BUSINESS INTELLIGENCE ENGINEERING

MEMORY THEME:

Face recognition and classification using deep learning.

Realized By:

GAMOURA Abderrazak

KAMMAR Amar

Supervised By:

Mme.BENABIDE Sonia

Members of jury:

Dr. ATTIA Abdelouahab

Academic Year: 2020/2021

Acknowledgment

We first of all thank Allah "the Almighty, for giving us courage and patience, and for guiding us to where we have arrived.

We would like to extend our sincere thanks to the people who have given us their assistance and who have contributed to the development of this brief.

We address a big thank you to the person in charge of this memory, Mme. Benabid Sonia for her precious help and for the time she devoted to us as well as for the help she gave us.

It is a pleasure for us as much as a duty to thank all the people who have been able to contribute directly or indirectly to the accomplishment of this project.

SUMMERY

Chapter 01: Biometric Systems

Chapter 02: Face recognition system

Chapter 03: Convolutional neural network

Chapter 04: Experimental and results

List of figures:

Abbreviations:

- CNN: Convolution Neural Network
- ConvNet: Convolution Neural Network
- AI: Artificial Intelligence
- ANN: Artificial Neural Network
- DNNs: Deep Neural Networks
- AE: Autoencoder
- DBN: Deep Belief Network
- GAN: Generative Adversarial Network
- DBM: Deep Boltzmann Machine
- SOM: Self- Organizing Map
- RNN: Recurrent Neural Network
- RBFN: Radial Basis Function Network
- DL: Deep Learning
- SVM: Support Vector Machine
- LBP: Local Binary Pattern
- PCA: Principal Component Analysis
- LDA: Linear Discriminant Analysis
- PC: Principal Components

Abstract

Biometrics is the automatic identification of a person based on their physiological or behavioral characteristics, such as fingerprints, voice and face. In this case, facial recognition is currently a booming field.

It is gradually entering our lives through our mobile phones or our computers portable, also used in the service of companies and in security and management.
The work proposed in this thesis aims to develop or characterize an algorithm offering expertise in this particular biometric field, and allow, more less, to facilitate its application in the management of attendance.

Despite the improvement in the detection rate, it is currently the subject of numerous studies. The objective of our project will be to implement a system of identification of people, this identification is based on images. These sequences are analyzed and we extract the visual information related to the face, we then build our database and we present the technique to recognize a person among the set. The facial recognition system is based on the CNN principle and we use the MATLAB library and its implementation This detection will assist the operator and then allow him to perform various tasks in the management of the presence of his staff (Presence, Absence, Delays, etc.).

Résumé

La biométrie est l'identification automatique d'une personne en fonction de ses caractéristiques physiologiques ou comportementales, telles que les empreintes digitales, la voix et le visage. Dans ce cas, la reconnaissance faciale est actuellement un domaine en plein essor.

Il entre progressivement dans nos vies à travers nos téléphones portables ou nos ordinateurs portables, également utilisés au service des entreprises et dans la sécurité et la gestion.

Le travail proposé dans cette thèse vise à développer ou caractériser un algorithme offrant une expertise dans ce domaine biométrique particulier, et permettre, plus moins, de faciliter son application dans la gestion des présences.

Malgré l'amélioration du taux de détection, elle fait actuellement l'objet de nombreuses études. L'objectif de notre projet sera de mettre en place un système d'identification des personnes, cette identification est basée sur des images. Ces séquences sont analysées et nous extrayons les informations visuelles liées au visage, nous construisons ensuite notre base de données et nous présentons la technique pour reconnaître une personne parmi l'ensemble. Le système de reconnaissance faciale est basé sur le principe CNN et nous utilisons la librairie MATLAB et sa mise en œuvre Cette détection va assister l'opérateur et lui permettre ensuite d'effectuer diverses tâches dans la gestion de la présence de son personnel (Présence, Absence, Retards, etc..).

الملخص

القياسات الحيوية هي التحديد التلقائي للشخص بناءً على خصائصه الفسيولوجية أو السلوكية ، مثل بصمات الأصابع ...والصوت والوجه. في هذه الحالة ، يعد التعرف على الوجه مجالًا مزدهرًا حاليًا.

إنها تدخل حياتنا بشكل تدريجي من خلال هواتفنا المحمولة أو أجهزة الكمبيوتر المحمولة لدينا ، كما تستخدم في خدمة الشركات وفي الأمن والإدارة يهدف العمل المقترح في هذه الأطروحة إلى تطوير أو تمييز خوارزمية تقدم الخبرة في هذا المجال البيومتري المحدد ، وتسمح ، بدرجة أقل ، بتسهيل تطبيقها في إدارة الحضور.

على الرغم من التحسن في معدل الكشف ، إلا أنه يخضع حاليًا للعديد من الدراسات. الهدف من مشروعنا هو تطبيق نظام لتحديد هوية الأشخاص ، وهذا التعريف يعتمد على الصور. يتم تحليل هذه التسلسلات ونستخرج المعلومات المرئية المتعلقة بالوجه ، ثم نبني قاعدة البيانات الخاصة بنا ونقدم التقنية للتعرف على شخص من بين المجموعة. يعتمد نظام التعرف على الوجه على مبدأ CNN ونستخدم مكتبة MATLABوتنفيذها. هذا الكشف سيساعد المشغل ومن ثم يسمح له بأداء مهام مختلفة في إدارة تواجد موظفيه (التواجد ، الغياب ، التأخير ، إلخ. .).

general introduction

In the current context, the security of information systems has become a very important research field, in particular, designing a reliable, efficient and robust identification system is a priority task. The identification of the individual has become essential to ensure the security of systems and organizations, Faced with this growing demand. several biometric recognition methods have been proposed, speaker recognition, facial recognition, fingerprint. recognition of the iris, the retina, the shape of the hand.

These methods have reached their limits, in terms of performance. For example, face or voice recognition are very well accepted by users but the rate of good facial identification is at best around 85%, which makes them too unsatisfactory for real applications.

Other methods are more reliable such as recognition of the retina or the iris, they are expensive and in general, poorly accepted by the general public. In addition, systems that rely on a single biometric modality are vulnerable to attacks. For the moment, no biometric indicator is 100% reliable according to. What gave birth to the fusion of multi modal biometric indicators all the arguments cited above plus the results of the various works have shown the performance of multi modal Biometric systems compared to Uni modal systems is a strong reason that led us to work on this subject.

The purpose of our work will be about how to implement a system of identification of people, this identification is based on images. These sequences are analyzed and extracted Visual information related to the face, so we built our database and presented the technology to recognize a person among all. Using the convolution neural network (CNN)

In the first chapter we present the biometric systems “voice. Iris. Geometric of the hand. Facial …” we put the definition of each system along with the advantages of each one in the second one we talked about the face recognition system. the methods used in it (methods of detection, feature extraction and calcification) along with different element that disturb the face recognition process then we give the advantages and the disadvantages of FR system in the third chapter we give some details about the convolutional neural network (architecture, layers.) with an explanation of how it works.

Finally using MATLAB, we experiment the CNN method in face recognition using the Resnet 18 training function with CNN and SVM classifier with yale faces dataset

Chapter 01

Biometric Systems

1. Introduction

Biometrics consists of identifying a person from one or more physiological characteristics (fingerprints, face, iris, hand contour, etc.), or behavioral (signature, gait, etc.). Another definition of biometrics is given by Roethenbaugh: "Biometrics applies to unique and measurable human characteristics or characteristics that automatically recognize or verify identity." Biometric systems have been increasingly used in recent years. The advent of the computer and its ability to process and store data has enabled the creation of computerised biometric systems.

Figure 1. 1. Biometric characteristics: a) DNA, b) Ear, c) face, d) infrared face) thermo gramme hand, f) hand vein, g) Fingerprints, h) walking, i) gesture j) iris, k) palm print, l) retina, m) signature, n) voice.

In this chapter, we will introduce some basic concepts and definitions related to biometrics. We will give the principle of operation of biometric systems as well as the tools used to measure their performance. We will especially emphasize the place of facial recognition among other biometric techniques, because it is the objective of this memory. Finally, we will expose the major difficulties related to face recognition, and which are still the subject of research by the scientific community.

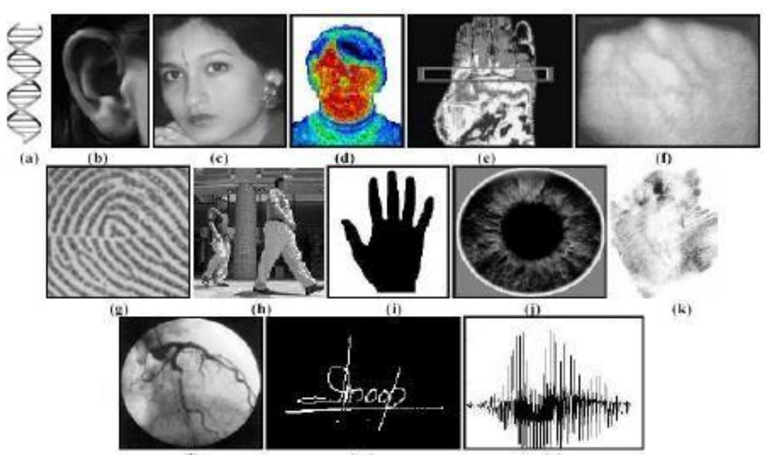


Figure 1. 1. Biometric characteristics

2. Presentation of some biometric technologies

No single biometric could effectively meet the needs of all identification applications. A number of biometric techniques have been proposed, analyzed and evaluated. Each biometric has its strengths and limitations, and as a result, each biometric is used in a particular application. For physical characteristics, we will describe face recognition, fingerprints, hand geometry and iris. For behavioral characteristics, we will describe biometrics based on voice and signature.

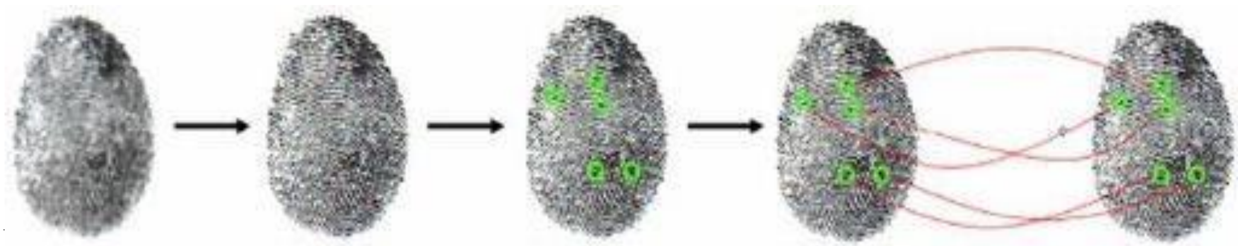
There are other biometric methods based on the veins of the hand, the D.N.A, body odor, ear shape, lip shape, keyboard typing rhythm, gait, which will not be developed in this chapter. [1]

2.1- Fingerprints

At present, fingerprint recognition is the most widely used biometric method. Fingerprints are composed of locally parallel lines with singular points (minutiae) and constitute a unique, universal and permanent pattern. To obtain an image of the fingerprint, technological advances are allowed to automate the task by means of integrated sensors, thus replacing the classic use of ink and paper. These sensors operate according to different measurement mechanisms (pressure, electric field, temperature) make it possible to measure the imprint of a fixed finger positioned on it (matrix sensor) or in motion (scanning sensors).

The fingerprint image of an individual is captured using a fingerprint reader and then the characteristics are extracted from the image and then a template is created. If proper precautions are followed, the result is a very accurate means of authentication.

Fingerprint matching techniques can be classified into two categories: techniques based on local minutiae detection and techniques based on correlation. The minutiae-based approach is to first find the minutiae points and then trace their locations on the finger image



This method does not take into account the overall structure of ridges and furrows.

Correlation-based methods are able to overcome the problems of the minutiae-based approach. These methods use the overall structure of the footprint, but the results are less accurate than with minutiae. In addition, correlation techniques are affected by the translation and rotation of the image of the footprint. This is why the two approaches are usually combined to increase system performance. [1]

2.2- The voice

Of all the human traits used in biometrics, the voice is the one that humans learn to recognize from an early age. Speaker recognition systems can be divided into two categories: pronounced text-dependent systems and text-independent systems. In the first case, the user is required to use text (a word or phrase) fixed predetermined during learning and recognition sessions. Whereas, for a text-independent system the speaker speaks freely without predefined text.

The latter category is more difficult, but it is useful in the event that one needs to recognize a speaker without his cooperation. Research on speaker recognition is growing, as it does not require expensive hardware, since most personal computers nowadays are equipped with a microphone. However, poor quality and ambient noise can influence verification and consequently reduce its use in biometric systems. In a speaker recognition system, the signal is first measured and then broken down into several channels of band-pass frequencies.

Then, the important characteristics of the voice signal are extracted from each band.

Among the most commonly used characteristics are cestreaux coefficients. They are obtained by the logarithm of the Fourier transform of the vocal signal in each band. Finally, the mapping of cestreaux coefficients makes it possible to recognize the voice. In this step, approaches based on hidden Markov models, vector quantization, or dynamic time distortion are usually used.

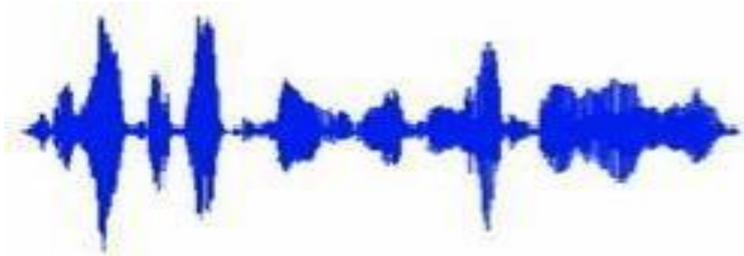


Figure 1.3: Spectrum of a voice signal

2.3- The iris

The use of the iris as a unique biometric characteristic of humans has resulted in reliable and extremely accurate identification technology. The iris is the region, in the form of a ring, located between the pupil and the white of the eye, it is unique. The iris has an extraordinary structure and offers many texture characteristics that are unique for each individual. The algorithms used in iris recognition are so accurate that the entire planet could be registered in an iris database with few identification errors. The iris image is usually captured using a standard camera. However, this stage of capture involves the cooperation of the individual. In addition, there are several constraints related to the use of this technology. For example, it is necessary to ensure that the individual's iris is at a fixed distance and close to the capture device, which limits the use of this technology.

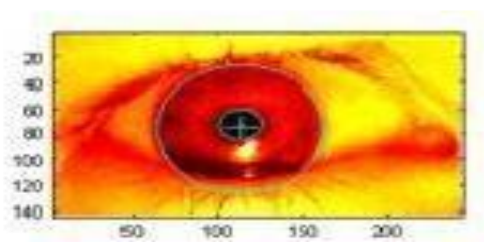


Figure 1.4: Iris photo

2.4- The geometry of the hand

Hand geometry is a recent biometric technology. As the name suggests, it consists of analyzing and measuring the shape of the hand, that is, measuring the length, width and height of a user's hand and creating a 3-D image. some Infrared and a digital camera are used to acquire data from the hand.

This technology offers a reasonable level of accuracy and is relatively easy to use.

However, she can be easily deceived by twins or by people with close hand shapes. The most popular uses of hand geometry include presence recording and access control. On the other hand, hand geometry capture systems are relatively large and cumbersome, which limits their use in other applications such as authentication in embedded systems:
Mobile phones, cars, laptops, etc.

Figure 1.5: Hand geometry recognition device

2-5 The face

Our faces are complex objects with features that can vary over time. However, humans have a natural ability to recognize faces and identify people at a glance. Of course, our natural recognition capability extends beyond face recognition, where we are also able to quickly spot objects, sounds or smells. Unfortunately, this natural ability does not exist in computers. This is how the need to artificially simulate recognition was born in order to create autonomous intelligent systems. Simulating our natural ability to recognize faces in machines is a difficult but not impossible task. Throughout our lives, many faces are seen and kept naturally in our memories forming a kind of database. [1]

Figure 1.6. The ten views of a person in the ENT database

Conclusion

In this chapter, we have presented the deferent types of biometrics. This study allowed us to see technologies used in biometric systems for the identification of persons.

Chapter 02

Face recognition systems

I. introduction

For years Facial recognition has been a subject of research in the field of computer vision. Now facial recognition has become a reality, facial recognition systems are the most common and popular. Facial recognition is a biometric identification technique based on automatic processing of digital images of an individual and allowing him to be identified from the characteristics of his face. Inspired by the workings of the human eye. The principle is simple: a sensor "captures" a face, transforms it into digital data and then compares it to a database, the latter two operations being performed by an algorithm. Facial recognition makes it possible to adapt biometric verification to all situations, It is a very effective technology that is used in many security-related applications. For example, it is a very reliable tool to help police forces identify criminals, or to allow customs services to verify the identity of travelers. Currently, with the digitization of exchanges, the use of this technology is spreading to the business world. Used in commercial applications, facial recognition makes it possible, for example, to secure online transactions. Facial recognition is contactless and its use requires no specific tools, making it the ideal solution for identifying people in a crowd or in public spaces.

Facial recognition is an ability that links a person's appearance to their identity. During a meeting, this skill makes it possible to remember previous exchanges and thus to build a long-term relationship where individuals end up knowing each other and knowing how to anticipate their respective behaviors and needs. Facial recognition is therefore a technology that has reached a certain maturity to the point of making it a tool not only of everyday life but also a solution among others to improve security. The challenge is that of technical standards that guarantee a level of quality and wide exploitation. It is also an economic issue for the development of products and their exploitation.

2/ Presentation of the facial recognition system

Facial recognition systems are automated systems capable of identifying individuals based on the characteristics of their face such as the spacing of the eyes, edges of the nose, the commissures of the lips, ears, chin, etc. These characteristics are analyzed and then compared to an existing database in order to identify a person or verify their identity. In a facial recognition system, an image follows a very specific process since its entry to determine the identity of the face wearer. This process consists of several steps that can be illustrated by the following diagram:[2]

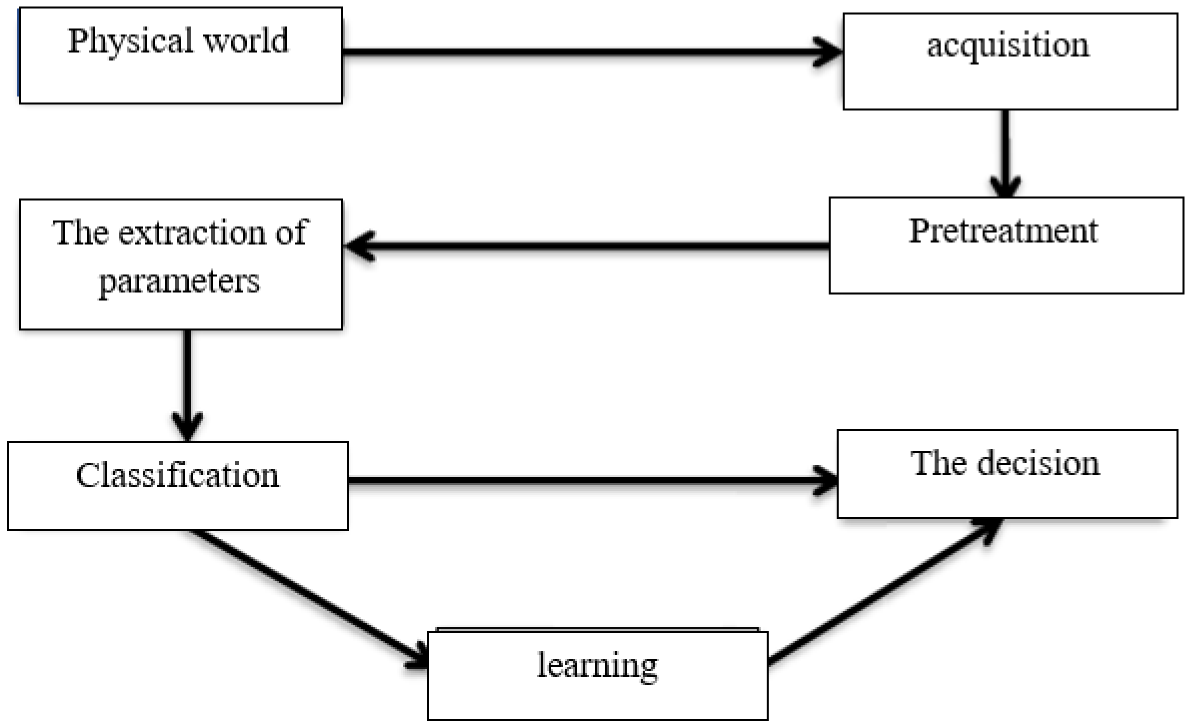


Figure 2. 1. Process of a facial recognition system.

So, to be identified, the image of a person in a recognition Facial system follows the following process:

2.1 Facial recognition process:

2.1.1 The physical world:

This is the real world outside the system before the image is acquired. In this Step, we generally take into account three essential parameters: Lighting, variation in posture and scale. The variation of any of these three parameters can drive at a distance between two images of the same individual, greater than that separating two images of two different individuals, and consequently a false identification.

2.1.2 Image acquisition:

This is the operation that makes it possible to extract from the real world a matrix representation i.e., an image, this operation can be static (Camera, Scanner, etc.) or dynamic (Camera, Web Cam).[2]

2.1.3 Pretreatments:

The raw data from the sensors are the initial representations of the data, from which of the treatments make it possible to build those that will be used for recognition. The raw image can be affected by different factors, thus causing its deterioration, it can be noisy, i.e., containing parasitic information because of optical or electronic devices. To compensate for these problems, there are several methods of processing and enhancing images, such as: normalization, histogram equalization, etc. ....

2.1.4 The extraction of parameters:

In addition to classification, the parameter extraction step represents the heart of the recognition system, it consists in carrying out the image processing in another workspace that is simpler and that ensures a better exploitation of data, and therefore allow the use, only, of useful, discriminating and non-redundant information.

2.1.4.1 Characteristic extraction algorithms:

a/ FAST algorithm:

Features from accelerated segment test (FAST), which can be translated as characteristics from accelerated segment tests, is a feature detection algorithm, presented by researchers at the University of Cambridge, used in the field of computer vision for object detection. The algorithm selects a pixel "p" and the appropriate threshold value "t", then consider a circle of 16 pixels around the pixel being tested, if the set of "n" pixels in the circle of 16 pixels are all brighter than "Ip+t" or darker than "Ip-t" then the point of interest is a corner.

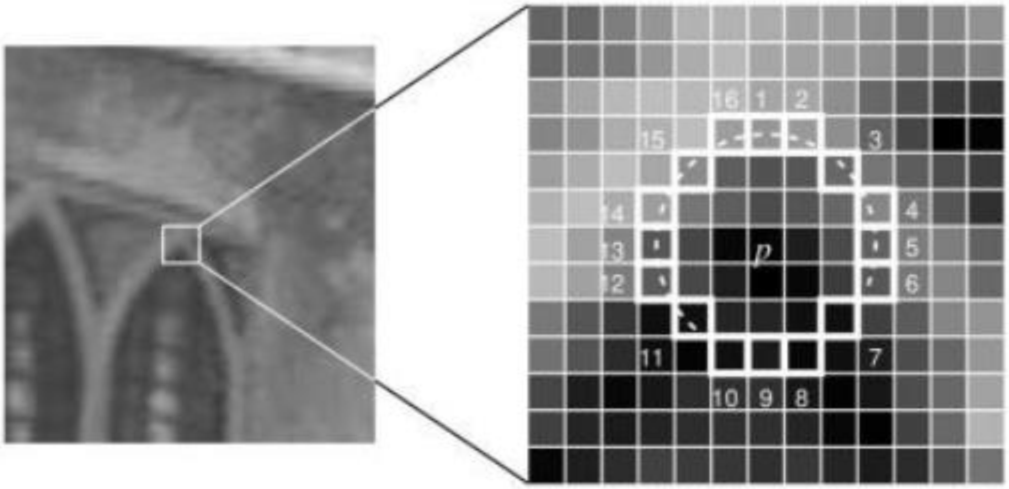


Figure 2. 3. FAST Algorithm

• **Advantages:**

- high performance.

• Disadvantage:

- It does not reject as many candidates for <12.

- The choice of pixels is not optimal, because its effectiveness depends on the order of the questions and the distribution of corner appearances.

- The results of high-speed tests are discarded.

- Several characteristics are detected adjacent to each other.[3]

b/ Harris algorithm:

Here's how to detect features in a 2D image, using the so-called Des Harris corners method. Harris' idea is as follows: at the level of these characteristic pixels (which he calls "Corners": corners, angles), the intensity of the image will vary significantly in several directions.

So, we're going to look at the variation in the intensity of the image around each pixel.

Definition 1 (Harris matrix). Let p be a pixel of the image. Let V be a neighborhood around p in the image. for any pixel q in the image, Iq is the intensity at pixel q. The Harris matrix for the pixel p is the Matrix

$$H_p = \sum_{q \in V} w_q \nabla I_q (\nabla I_q)^t$$

With wq a weight associated with q and ∇ the gradient operator. Wq weights must be circularly symmetrical (i.e., all pixels at the same distance of p must have the same weight). Hp is therefore a 2 × 2 matrix. The wq weights associated with pixels in the vicinity of p are usually chosen in such a way as to form a Gaussian mask 3 × 3 or 5 × 5: if q − p has for coordinates (x, y)

$$w_q = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

Harris' detection algorithm is as follows:

1. Set T a threshold.

2. For any pixel p of the image:

(a) calculate Hp

(b) calculate $c(H_p) = \frac{determinant(H_p)}{trace(H_p)}$.

3. Keep p pixels such as c(Hp) > T.

4. Of these pixels, keeping those such as C(Hp) is a local maximum in a 3×3 neighborhood (at least) The result pixels will form our feature set.

Benefits:

o The Harris corner detector is invariant to translation, rotation and lighting change, this detector is the most repetitive and informative.

• Disadvantage:

o it is not invariant to large-scale change.

c/ Canny algorithm:

The Canny Edge Detector technique is very important for detecting the edges of an image. This operator isolates the noise from an image before finding the edges of an image without affecting the characteristics of the image, and then applies the tendency to find the edges and the critical value for the threshold.[3]

Figure 2. 4. Canny algorithm

• **Advantages**:

o Use probability to find the error rate.

o Location and response.

o Improved signal/noise ratio.

o Better detection especially in noise conditions.

• Disadvantage:

o Complex calculations.

o False zeroing problem.

o A lot of time.

d/ Roberts Algorithm:

Is a two-dimensional spatial gradient measure on an image. It highlights regions of high spatial frequency that often correspond to edges. The values of

pixels at each point of output represent the estimated absolute amplitude of the spatial gradient of the input image at that point.

+1	0
0	-1

0	+1
-1	0

• **Advantage**:

o A simple, quick to calculate.

• **Disadvantages:**

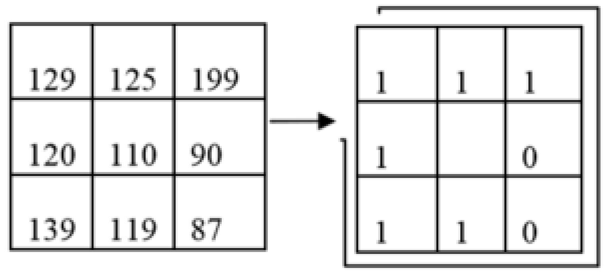
o High sensitivity to noise.

o Few pixels are used to approximate the gradient.

e/ LBP Algorithm (Local Binary Pattern):

The local binary pattern descriptor (LBP) is an effective method of extracting

Facial features; It extracts information from neighboring pixel values and expands the histogram of the image. It is a nonparametric operator and describes the local spatial structure of an image. It calculates a binary code from binary derivatives of pixels and then finds the difference of the central pixel with its neighboring pixels, arranges these differences of an ordered shape and finally this pattern of bits is converted to decimal value which is the new LBP code for the central Pixel. The basic LBP operator works for 3 × 3 pixels is described by the following diagram:[3]



Binary Code: 11100111

LBP Code: 231

The figure above shows grayscale values of 3 × 3 pixels and the LBP code is calculated using the following formula:

$$LBP(x_c, y_c) = \sum_{n=0}^{7} S(i_n - i_c)2^n$$

$$S(x) = \begin{cases} 1 \; if \,(x \geq 0) \\ 0 \; if \,(x < 0) \end{cases}$$

Here xc and yc show the position of the central pixel, in and ic are grayscale values of the surrounding pixels and the central pixel respectively. After labeling the image with LBP codes, histogram image is generated that allows to recognize micro-patterns in the image such as eyes, nose and lips... etc.

• **Advantage:**

o Tolerance for changes in enlightenment and simplicity.

• **Disadvantage:**

o the operator produces a fairly long Histogram and is not too robust on the flat image.

f/ GABOR algorithm:

GABOR filters are applied to the image to extract multiple scales or frequencies aligned at different angles. The characteristics are selected to represent a single biometric in the image.[3]

GABOR filter of orientation θ and wavelength ʎ and extension σ:

$$g(\sigma,\theta,\lambda,I,i,j)=\frac{1}{2\pi\,\sigma^2}\sum_{k,l} e^{-(\frac{k^2+l^2}{\sigma^2})}.e^{2\pi\,i(\frac{k.\cos\theta+l\sin\theta}{\lambda})}I(i+k,j+l)$$

L'énergie de l'image via ce filtre :

$$E_g(\sigma,\theta,\lambda,I)=\frac{1}{N}\sum_{i,j}|g(\sigma,\theta,\lambda,I,i,j)|^2$$

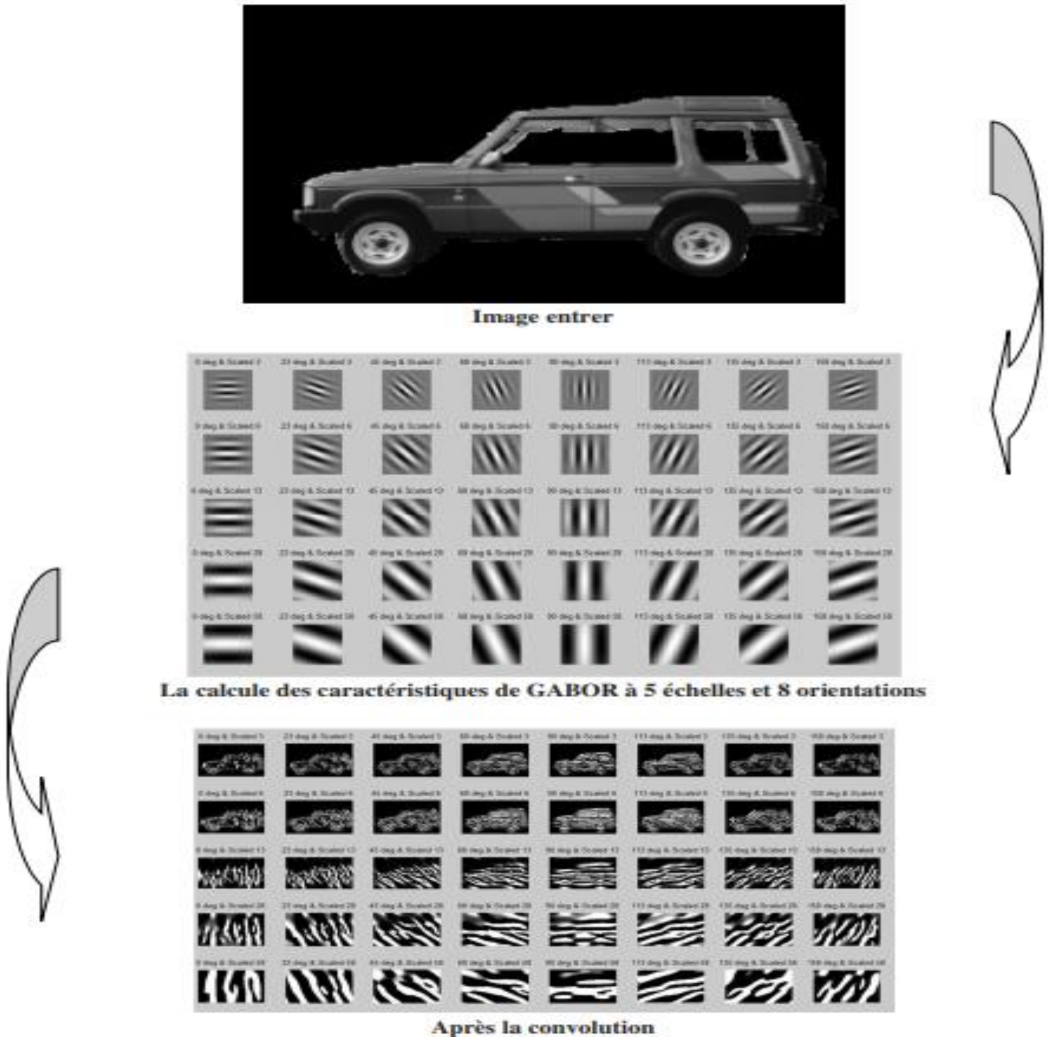


Figure 2. 5. GABOR Algorithm

Advantage:

• GABOR filters produce superior results for segmentation of

textures.

• Rotating invariant.

• Better segmentation results.

Disadvantage:

• Very heavy algorithm.

• Response time is very slow.

Remark:

Table below shows a comparison between the algorithms for extracting

Characteristics:

	type de caractéristique	rotation	bruit	luminosité	échelle
FAST	Coin	mauvais	moyen	mauvais	moyen
Harris	Coin	mauvais	moyen	moyen	mauvais
Canny	Bords	Moyen	Très bien	Bien	mauvais
Roberts	Bords	Moyen	mauvais	Bien	mauvais
Sobel	Bords	Moyen	mauvais	Bien	mauvais
GABOR	Bords/texture	Trés bien	Bien	Très Bien	Très bien
LBP	texture	Tres bien	mauvais	Très bien	mauvais

Table 2.1: Deference between extraction algorithms

For nearly three decades, the use of features based on GABOR filters has been promoted for their useful properties in image processing. The most important properties are related to invariance to illumination, rotation, scale and noise. As we see in the table, GABOR is better than the rest of the algorithms, these properties are based on the fact that they are all parameters of the GABOR filters themselves, the latter has been used in many computers vision tasks, such as texture segmentation, face detection and iris recognition which is why I chose this algorithm to be used in my project.[4]

2.1.5 Classification: (modelling)

This step involves modeling the parameters extracted from an individual's face or set of faces based on their common characteristics. A model is a set of useful, discriminating, non-redundant information that characterizes one or more individuals with similarities.

2.1.6 Learning:

This is the stage where individuals are taught to the system, it consists in memorizing the parameters, after extraction and classification, in a well-ordered database to facilitate the recognition phase and the making of a decision, it is in a way the memory of the system.[5]

2.1.7 The decision:

The decision is the part of the system where we decide whether an individual belongs to all faces or not, and if so, what is his identity. So, the decision is the culmination of the process. It can be valued by recognition rate (reliability) which is determined by the rate of accuracy of the decision.

2.2 Issue:

The problem of face recognition can be formulated as follows: being given one or more images of a face, the task is to find or verify the identity of a person by comparing his face to all the face images stored in a database. In this work, we limit ourselves to recognition from a 2D image of a face in unstrained environments. Such systems must be able to overcome the following problems:

2.2.1 Influence of variations in installation:

Changes in orientation and changes in the angle of inclination of the face cause many changes in appearance in the collected images. Deep rotations cause the occultation of certain parts of the face as for three-quarters views. On the other hand, they bring differences in depth that are projected onto the 2D plane of the image, causing deformations that vary the overall shape of the face. These deformities, which correspond to the stretching of certain parts of the face and the compression of other regions, also vary the distances between facial features.

2.2.2 Influence of lighting changes:

The intensity and direction of lighting when shooting greatly influences the appearance of the face in the image. Indeed, in most common applications, changes in lighting conditions are inevitable, especially when views are collected at different times, indoors or outdoors. Given the specific shape of a human face, these variations in lighting can reveal shadows accentuating or masking certain facial features.

2.2.3 Influence of facial expressions:

Faces are non-rigid elements. Emotional facial expressions, combined with speech-induced distortions, can produce significant changes in appearance, and the number of possible configurations is too large for them to be described in extenso realistically.

2.2.4 Influence of occultations:

A face can be partially masked by objects or by wearing accessories such as glasses, a hat, a scarf. Occultations can be intentional or not. In the context of video surveillance, this may be a deliberate desire to prevent recognition. It is clear that recognition will be all the more difficult as few discriminating elements are simultaneously visible.

2.3 Fields of application:

Facial recognition can be found in several areas:

a-information security: database security, file encryption, intranet security, internet access, register on a personal installation.

b-right of access and ssurveillance: Advanced video surveillance, access control, event analysis, prosecution of suspects and investigation.

c-Security: in stadiums, airports and shopping centers in several countries to prohibit the access of certain individuals on file.[5]

2.4 Advantages and disadvantages of facial recognition

advantages	Disadvantages
Well accepted technology by public	Sensible technology to environment
Recognition rate is much higher in a good lighting and a fixed position	Sensible technology to changing such as mustache, new haircut

Table 2. 2.Advantage and disadvantage of Facial Recognition

2.5 Face detection methods:

Face detection in the image is an indispensable and crucial treatment before the recognition phase. Indeed, the face recognition process can never become fully automatic if it has not been preceded by an effective detection step. Processing consists of looking in an image for the position of the faces and extracting them in the form of a set of thumbnails in order to facilitate their further processing. A face is considered correctly detected if the extracted thumbnail size does not exceed 20% of the actual size of the facial region, and it contains essentially the eyes, nose and mouth. The interest of facial location goes beyond the application of this brief. Its usefulness manifests itself in various fields ranging from video surveillance to interactive gaming. The first difficulties encountered by methods working to detect faces are variations in the pose of expression, facial rotation, age and illumination. For the rest, the difficulty is all the greater as most applications using this technology require real-time execution, limiting the algorithm's room for maneuver. The methods are divided into four categories. These categories can overlap if an algorithm can belong to two or more categories. This classification can be done as follows:

2.5.1 Knowledge-based methods:

These methods are based on the knowledge of the different elements that make up a face and the relationships that exist between them. Thus, the relative positions of different key elements such as the mouth, nose and eyes are measured to then serve as a

the classification 'face' 'not face' in Chiang et al. The problem with this type of method is that it is difficult to uniquely define a face. If the definition is too detailed, some faces will be missed while if the description is too general, the false positive rate will skyrocket.

2.5.1.1 Invariant approaches to characteristics:

These approaches use elements that invariant with variations in illumination, orientation, or expression such as texture or skin color signature for detection.

2.5.2 Model matching methods:

Templates can be defined either "manually" or parameterized using functions. The idea is to calculate the correlation between the candidate image and the template. These methods still encounter some problems of robustness related to variations in light, scale, etc. invariant to changes in brightness to characterize the different parts of the face.

2.5.3 Appearance-based methods:

These approaches typically apply machine learning techniques. Thus, the models are learned from a set of images representative of the variability of the appearance of the face. These models are then used for detection. These methods have the advantage of being executed very quickly but require a long training time. Methods in this category have shown good results compared to the other three types of methods. These include the method based on the neural networks of

Rowley et al, the method of **Schneiderman and Kanade** based on a naïve Bayes classifier and the famous algorithm of **Viola and Jones** operating in real time.

2.6 Face recognition methods:

Face recognition systems are very often ranked from the findings of psychological studies on how men use

2.6.1 Global Methods:

These so-called global or holistic methods are methods that use the entire region of the face as input information. Its main disadvantage may lie in the size of the data to be stored during the training phase. However, with the current storage capacity of our computers, we can relatively overcome this problem.

Among the global methods, we can mention:

_ **PCA** Principal Component Analysis (also called EigenFace).

_ **Linear** Discriminant Analysis (also called FisherFace): Methods based on Linear Discriminant Analysis (LDA) determine the most discriminating projection directions in eigenspace. To do this, they maximize inter-person variations compared to intra-person variations. Thus, the ADL-based method is then reduced to the eigenface method.

_ The Probabilistic approach transforms the problem of face identification into a problem of classification into two classes. It evaluates the probability of the difference between a test image and a prototype image belonging to the intra-person and inter-person classes. Note that the intra-person distribution cannot be evaluated in the case of one example per person, and the method also comes down to the eigenface method.

2.6.1.1 PCA:

Although there are many facial recognition algorithms that work well in constrained environments. Various changes in images present a great challenge in the face of a recognition system that must be robust with regard to the great variabilities of facial images such as facial expressions, facial pose and lighting. To cope with this problem, it is important to choose an appropriate representation of facial images. This representation must be compact and meaningful as PCA The PCA algorithm, and also known as Eigenfaces since it uses eigenvectors and eigenvalues. (Respectively Eigenvectors and Eigenvalues). This algorithm is based on well-known statistical properties and uses linear algebra. It is relatively quick to implement but it is sensitive to the problems of exposure lighting and facial expression. It is the basis of many current global algorithms. The main idea is to express the M learning images according to a base of particular orthogonal vectors, containing information independent of one vector to another.

These new data are therefore expressed in a way that is more appropriate for facial recognition. We want to extract the characteristic information from a face image, to encode it as efficiently as possible in order to compare it to a database of similarly encoded models. In mathematical terms, this amounts to finding the eigenvectors of the covariance matrix formed by the different images of our learning base. An image is treated as a vector in a large vector space, by concatenation of columns

2.6.1.2 LDA

Linear discriminant analysis (LDA), can be drawn from an idea suggested by RA Fisher in 1936. When LDA is used to find the subspace representation of a set of face images, the resulting base vectors defining that space are known as Fisherfaces.

The LDA algorithm was born from the work of Belhumeur et al. From Yale University (USA), in 1997. It is also known as "Fisherfaces". Unlike the PCA algorithm, that of the LDA performs a real separation of classes. To be able to use it, it is therefore necessary to first organize the image learning base into several classes: one class per person and several images per class. The LDA analyzes the eigenvectors of the data dispersion matrix, with the aim of maximizing variations between images of different individuals (interclasses) while minimizing variations between images of the same individual (intraclasses).

However, when the number of individuals to be treated is lower than the resolution of the image, it is difficult to apply the LDA which can then reveal matrices of singular dispersions (not invertible). As the PCA does not take into account class discrimination but LDA resolve this issue, and that standard LDA-based methods such as Fisherfaces, first apply the PCA for dimension reduction and then discriminant analysis. Appropriate questions about the PCA are usually related to the number of principal components (PCs) used and how they affect performance.[6]

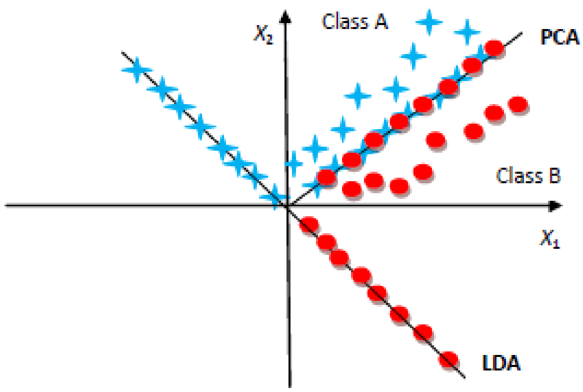


Figure 2.6 -Comparison between the projections of two classes of points ("class 1" and "class 2") on the main axes built by PCA and LDA.

2.6.2 Local Methods

This category is divided into two subcategories based respectively on local characteristics and appearances.

2.6.2.1 Methods based on local characteristics

Here geometric properties are extracted from the location of key points on the face. Two weaknesses arise directly from this: on the one hand the location of such points is not always an easy task when occlusions or variations in position or expression occur and on the other hand all the information necessary for robust recognition is not necessarily contained in these facial characteristics to recognize others. From this point of view, we distinguish the three categories:

global methods, local methods and hybrid methods. some key points, indeed a lot of information goes to the trap when the image is compressed to the information contained in a few places.

2.6.2.2 Method based on local appearances

Here it is several vectors corresponding to characteristics of the face that are used as input. These methods are a priori better suited for the single-sample problem [10]. First of all because a set of several vectors of small dimension instead of a single one of large dimension allows from the beginning to tackle the curse of dimensionality (curse of dimensionality). Then the fact of having several sub-vectors of characteristics allows the use of a weight system giving priority in

the final decision to the sub-vectors identified as being the most discriminative, which improves performance [11]. Finally, a high number of characteristic subsectors can increase the diversity of classifier types used through a team of classifiers and thus improve the performance of the global classifier.

The methods mentioned above do not explicitly consider the existing relationships between local characteristics. It is conceivable that the use of this information would be beneficial to the recognition system.

2.6.3 Hybrid Methods

Hybrid methods are approaches that combine global and local characteristics in order to improve the performance of face recognition. Indeed, local characteristics and global characteristics have quite different properties. It is hoped that their complementarity can be used to improve classification.

2.7 Deep learning-based methods

2.7.1 Convolutional Neural Networks

CNN or ConvNet for Convolutional Neural Networks is a type of acyclic artificial neurons (feed-forward), in which the connection pattern enters neurons is inspired by the visual cortex of animals. The neurons of this region of the brains are arranged so that they correspond to overlapping regions during the paving of the visual field. Their functioning is inspired by

biological processes, they consist of a multilayer stack of perceptron's, the purpose of which is to pre-treat small quantities of information. Convolutional neural networks have broad applications in image and video recognition, recommendation systems and processing of natural language.[7]

A convolutional neural network can have several tens or even hundreds of layers, each of which learns to identify different characteristics of an image. Filters are applied to each image used for learning to different resolutions, and the output of each convoluted image is used as input to the layer following. The first filters can be very simple features, as by example brightness or edges, then switch to more complex features that uniquely define the object. Like other neural networks, CNNs consist of an input layer, an output layer, and many hidden layers between these two layers (Figure 3.1).

.

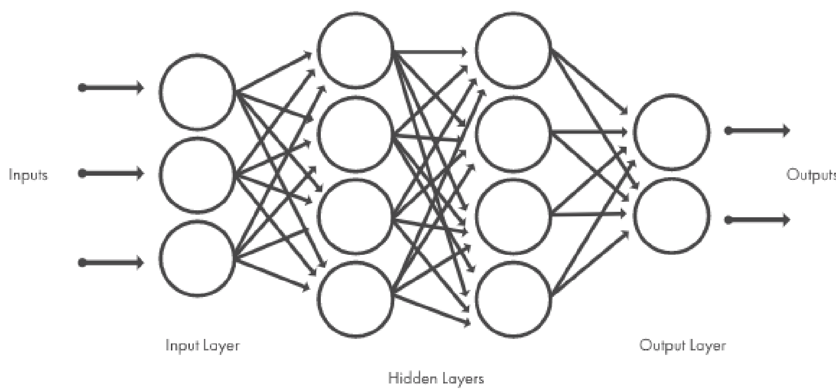


Figure 3.1- Overview of CNN Layers

All of these layers perform operations that modify the data in the lens learn the specific characteristics of this data. The three types of layers most common are: convolution, activation or RELU and pooling

Conclusion:

Face recognition technologies have been associated generally with very costly top secure applications. Certain applications of face recognition technology are now cost effective, reliable and highly accurate. As a result, there are no technological or financial barriers for stepping from the pilot project to widespread deployment. In this chapter we presented facial recognition methods and advantages of it.

Chapter 03

Convolutional neural network

1. Introduction:

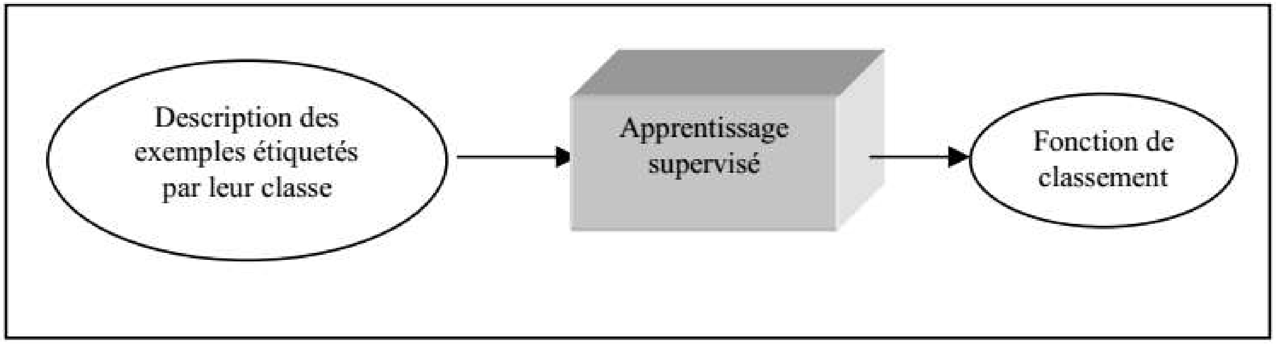
The world is filled with objects that almost all of us can understand and that make us react without thinking too much. For example, a "Stop" sign partly covered with snow remains a "Stop" sign, and a chair five times larger than a classic chair remains a place where to sit. However, for standard computers, this type of intuitive logic is out of the ordinary. scope. Today, machine learning is able to offer this advantage to computers thanks to its advanced technology. Machine learning is a sub-field of artificial intelligence (AI). In General, the goal of machine learning is to understand the structure of data and to integrate them into models that can be understood and used by the world. Although machine learning is a field of computer science, it differs from traditional IT approaches. Indeed, in the latter, the algorithms are explicitly programmed instruction sets used by computers to calculate or solve problems. Machine learning algorithms enable computers to train on data inputs and use statistical analysis to produce values that are within a specific range. For this reason, machine learning makes it easier to use computers in building models from sampling data to automate decision based on data entered in machine learning, tasks are usually classified into large categories. These categories are based on how learning is received or how the Feedback on learning is given to the developed system. Two of the methods the most widely adopted machine learning are supervised learning and unsupervised learning.

2. Machine learning:

Machine learning is a field of study of artificial intelligence that is based on mathematical and statistical approaches to give computers the ability to "learn" from data, i.e. improve their performance in solving tasks without being explicitly programmed for each. More broadly, it concerns the design, analysis, optimization, development and implementation of such methods. The concept of machine learning dates back to the mid-20th century. In the 1950s, the British mathematician Alan Turing imagines a machine capable of learning, a " Machine Learning". Over the following decades, different machine techniques Learning have been developed to create algorithms capable of learning and improve autonomously.[8]

2.1 Supervised learning:

In supervised learning, the computer is provided with examples of inputs that are labeled with the desired outputs. The purpose of this method is that the algorithm can "learn" by comparing



its actual output with the outputs "taught" to find errors and modify the template accordingly. Supervised learning therefore uses models for predicting label values on additional unlabeled data.

2.2 Unsupervised learning:

In unsupervised learning, the data is unlabeled, so that the learning algorithm alone finds commonalities among its input data. Because unlabeled data are more abundant than labeled data, the methods machine learning that facilitate unsupervised learning are particularly useful.

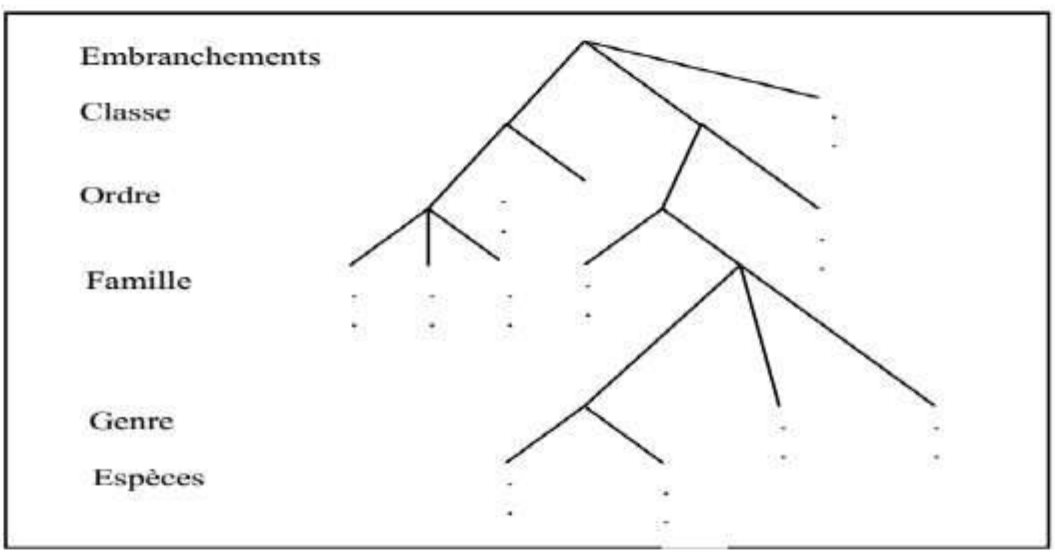


Figure 3.1-Excerpt from Linnaeus' taxonomic classification.

2.3 Deep learning:

2.3.1 Definition of deep learning:

Deep learning is a set of learning techniques automatic that has enabled significant advances in artificial intelligence in Years. In machine learning, a program analyzes a set of data in order to draw rules that will allow conclusions to be drawn about new data. Deep learning is based on what has been called, by analogy, "networks of artificial neurons", composed of thousands of units (the "neurons") that each perform small, simple operations. The results of a first layer of "neurons" serve input to the calculus of a second layer and so on. For example, for visual recognition, the first layers of units identify lines, curves, angles... upper layers identify shapes, combinations of shapes, objects, contexts... Advances in deep learning have been made possible in particular by the increase in the power of computers and the development of large databases ("big data").[9]

2.3.2 Fields of application of deep learning:

These techniques are developing in the field of computer science applied to ICT
(Visual recognition — e.g., of a traffic sign by a robot or an autonomous car (and voice car in particular) to robotics, bioinformatics, recognition or comparison of forms, safety, health, etc., assisted pedagogy by computer science, and more generally to artificial intelligence. Deep learning can, for example, allow a computer to recognize highly object objects deformable and/or analyze for example the emotions revealed by a photographed face or filmed, or analyze the movements and position of the fingers of a hand, which can be useful to translate sign language, improve the automatic positioning of a camera, etc... They are used for certain forms of medical diagnostic assistance (e.g.: automatic recognition of cancer in medical imaging), or prospective or prediction (e.g. prediction of the properties of a soil filmed by a robot).

3. Deep learning and convolution neuronal networks:

Artificial Neural Network (ANN) (Haykin, 2009) is a computational nonlinear model inspired by the biological systems in information processing. It consists of artificial neurons or processing elements and is typically organized in three types of interconnected layers. Data are presented to the network via the input layer, which communicates to one or more hidden layers where the actual processing is done via weighted connections. The hidden layers then link to an

output layer to give the output. It is possible to make the neural network more flexible and more powerful by using additional hidden layers. Artificial neural networks with many hidden layers between the input and output layers are called Deep Neural Networks (DNNs), and they can model complex relationships between the input and output. There are various deep neural networks used in face recognition.

Convolutional Neural Network (CNN) is the most popular. It shows outstanding results in image and speech applications. Autoencoder (AE) and its variants also gained much attention. They process data without using class labels and the purpose is to find patterns, such as latent subspaces. Generative Adversarial Network (GAN) has increased rapidly recently. It usually contains two nets, putting one against the other, thus called adversarial. It can learn to mimic any distribution of data. Deep Belief Network (DBN), Deep Boltzmann Machine (DBM) are also used in FR. However, the Recurrent Neural Network (RNN), Self- Organizing Map (SOM), Radial Basis Function Network (RBFN), are not used very often in FR.

3.1 The neural network:

The practice of all DL (Deep Learning) algorithms are neural networks. Neural Networks, also called ANN (Artificial Neural Networks), are models information processing that simulates the functioning of a nervous system biological. This is similar to how the brain manipulates information at the operation. All neural networks are made up of interconnected neurons that are organized in layers.[10]

3.1.1 The neuron:

What forms neural networks are artificial neurons inspired by the real thing.
neuron that is in our brain. The following 2 figures (Figure 3.1 and Figure 3.2)
show a representation of a real neuron and an artificial neuron.

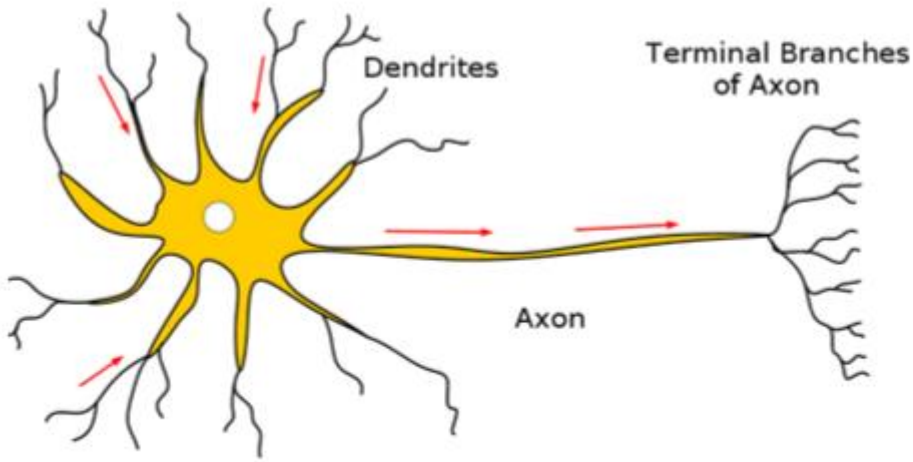


Figure 3.1: Representation of a neural network

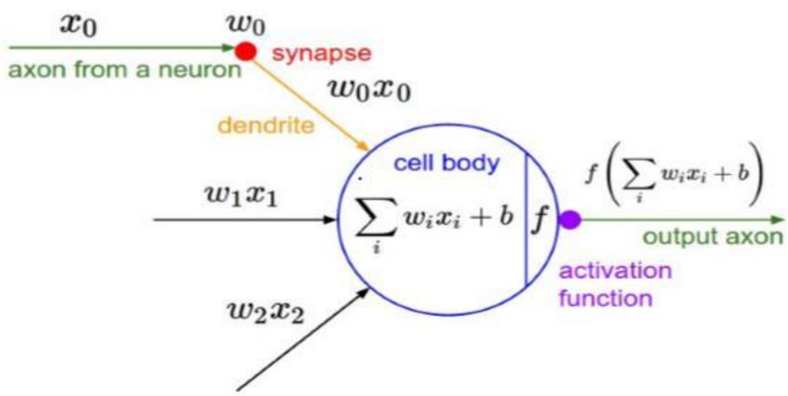


Figure 3.2: The neuron model

3.2 Neural network:

Neural Networks is one of the algorithms the most popular machine learning at the moment. Over time, it has been decisively proven that neural networks overpass others algorithms in terms of accuracy and speed. With various variants such as CNN (Convolutional Neural Networks (abbreviated as CNN)), RNN (Recurrent Neural network), Auto-Encoders, etc..., neural networks become little by little for scientists or machine learning practitioners, this that linear regression was for statisticians. Deep neural networks (CNNs) have had considerable success in recognition and localization of objects in images. The fundamental approach that led to the CNN is to build artificial systems based on the brain and human vision. Yet, in many important respects, THE CAPABILITIES OF THENS are inferior to anyone of human vision. A research axis promising consists in studying similarities and differences by filling in the gaps, to improve THENs. There are several Deep Learning algorithms.

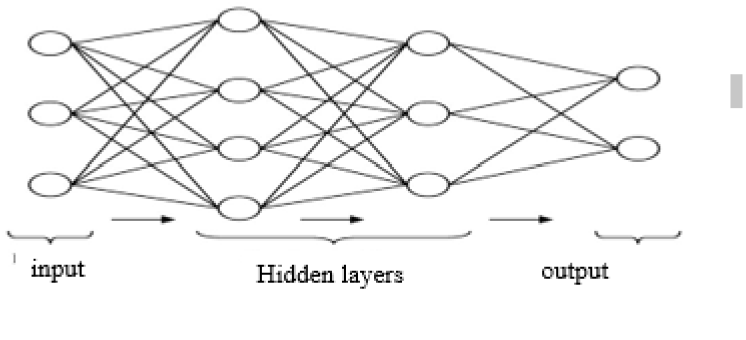


Figure 3. 3 Diagram of an artificial neural network

In our work, we are interested in reduction methods for Deep learning, it is essential to present the CNN which is fundamental to understand face recognition based on Deep Learning. In the following, we present an overview of the CNN.

3.2.1 Convolutional Neural Networks (CNN):

Computer vision is evolving rapidly by the day. One of the reasons is the development of deep learning. When we talking about vision by computer, a term convolutional neural network comes to mind because that CNN is heavily used here. Examples of CNN in computer vision are face recognition, image classification, etc. It is similar to the network of basic neurons. CNN also has parameters that can be learned, such as than the neural network, namely weightings, biases... An example of a schema in principle is illustrated in Figure 3.4. [11]

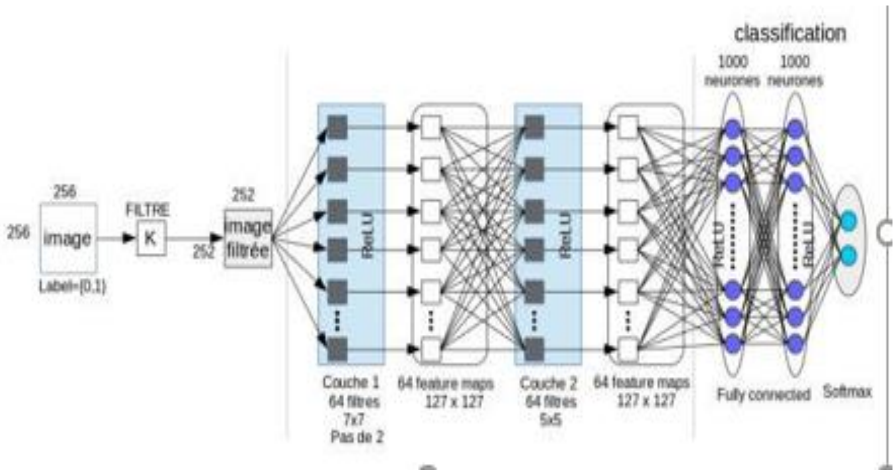


Figure 3. 4 Example of a CNN schema

3.2.2 Layers of convolutional neural networks:

There are several different layers in CNN as shown in Figure 1.5:

- Input layer.
- Convo layer (Convo layer: Convolution + ReLU).
- Pooling layer.
- Fully connected layer.
- Softmax/logistics layer.
- Output layer.

An example of the architecture of the CNN is shown in Figure 3.5

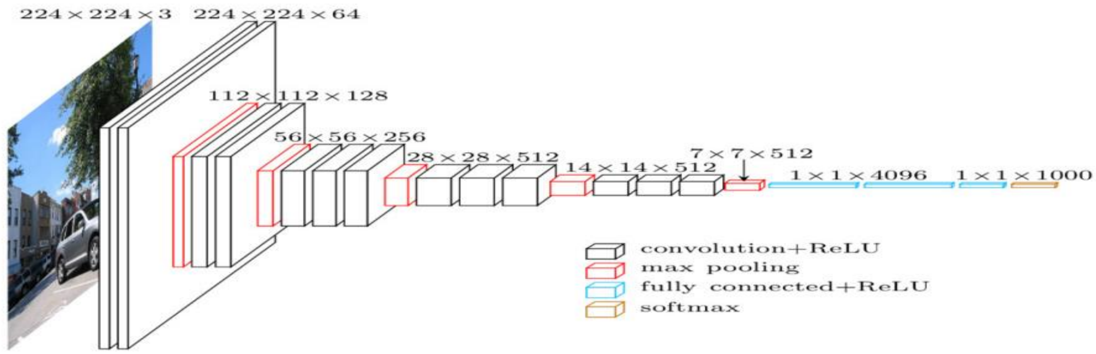


Figure 3. 5 Example of CNN architecture

CNN Input Layer:

The input layer in CNN must contain data describing the image. The Image data are represented by a three-dimensional matrix that in general must be reshaped into a single column (vector representation).

Convolution Layer:

The convolution layer is sometimes referred to as the extraction layer of characteristics, because the characteristics of the image are extracted in this layer. First, part of the image is connected to the Convo layer to perform a convolution operation and calculate the scalar product between the receiving field (this is a local region of the input image that is the same size as the filter) and the filter as shown in Figure 1.6. The result of the operation is a single integer of the volume output. Then we drag the filter to the next receiver field of the same image of entry by a stride and repeat the same operation. This operation is repeated by the same process over and over again until the entire image is Traveled. [12]

The formula to calculate tensor of the output of the feature maps of the layer is given by:

$$f^{(l)}[n,i,j] = b^{(l)}[n] + \sum_{c=1}^{C}\sum_{p=1}^{K}\sum_{q=1}^{K}\Phi^{(l)}[c,i+p,j+q].w^{(l)}[n,c,p,q]$$

• f (l) the tensor of the feature map output of the layer (l)

• b (l) [n] the bias applied to the feature n

• Φ (l) the identity tense ur of the feature maps of the layer (l)

• w (l) the pre-learned filter tensor

Padding:

Sometimes it can be interesting to keep a certain dimension in the sizes of the images out of convolutions. Padding is simply adding 0s all around a matrix (image) to increase its size. For example, padding 2 on a matrix size 32×32 will add 0s to the left of 2 columns, right on 2 columns, top on 2 rows and bottom on 2 rows!

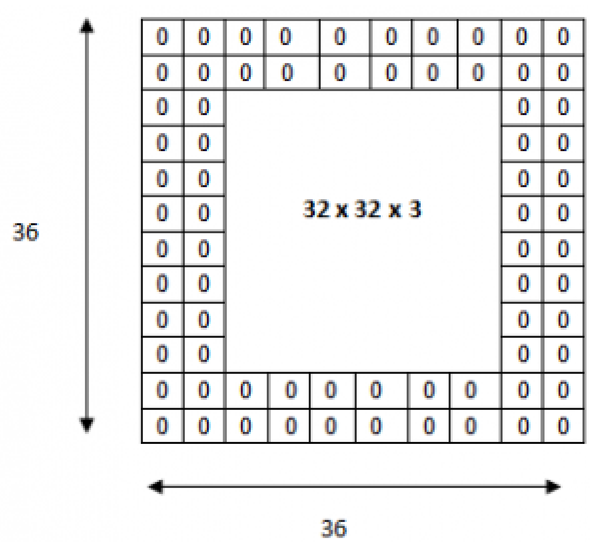


Figure 3.6: A padding of 2 on a matrix of size 32×32

The stride:

It controls the overlap of the receiving fields. The smaller the step, the more fields Receivers overlap and the larger the output volume will be.

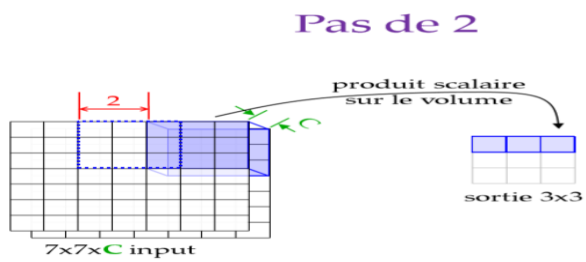


Figure 3.7: An example of a step of 2

The formula for calculating the number of neurons in the output volume is:

$$w_0 = \frac{w_i - k + 2P}{s} + 1$$

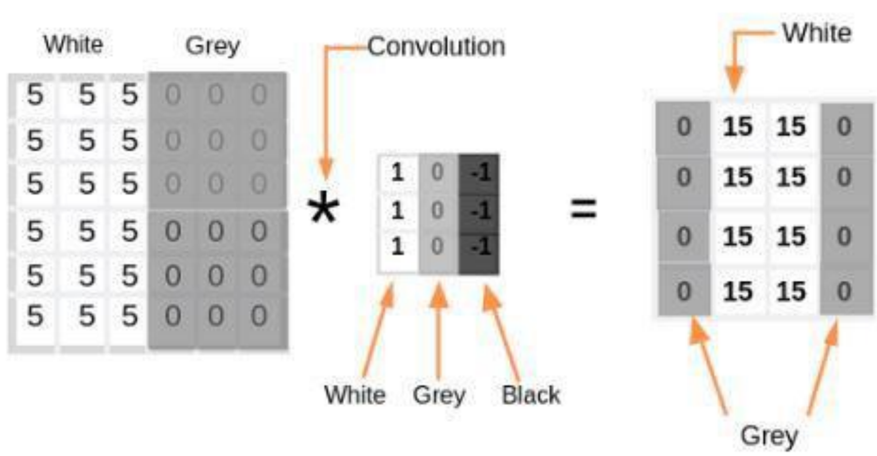


Figure 3.8 Example of the principle of the convolutional filter

The Convo layer also contains the ReLU activation see Figure 1.7 for that all negative values be set to zero...

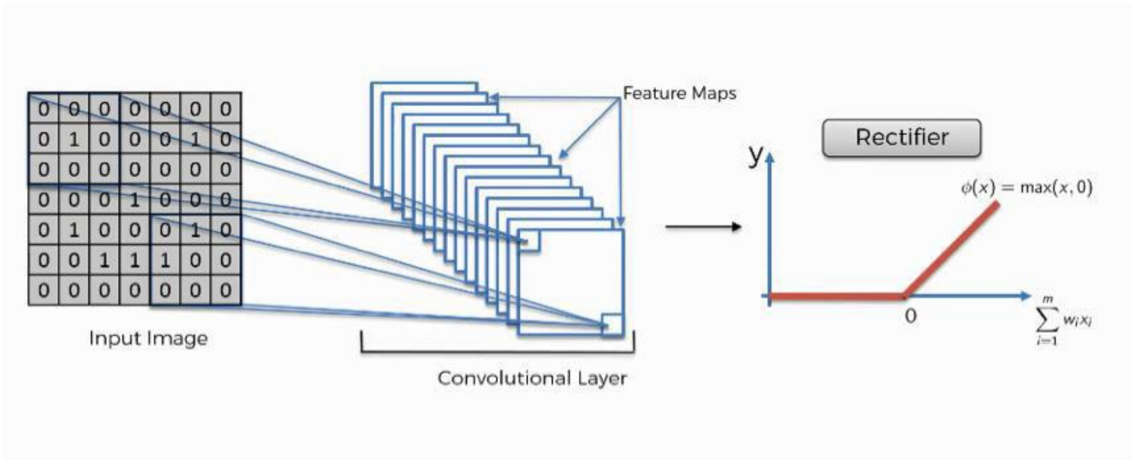


Figure 3.9 Principle of the ReRead function

Pooling Layer:

The Pooling layer is used to reduce the spatial volume of the image of entry after convolution. It is used between two layers of convolution. If we apply fc (Fully Connected) after the Convo layer without applying pooling or maximum pooling, the calculation will be expensive. Thus, the maximum common is the only way to reduce the spatial volume of the input image by encoding
information.

The function for the calculation is given by:

$$f^{(l)}[n,i,j] = \max_{p,q \in [1:K]} \left(\Phi^{(l)}[n, i+p, j+q] \right)$$

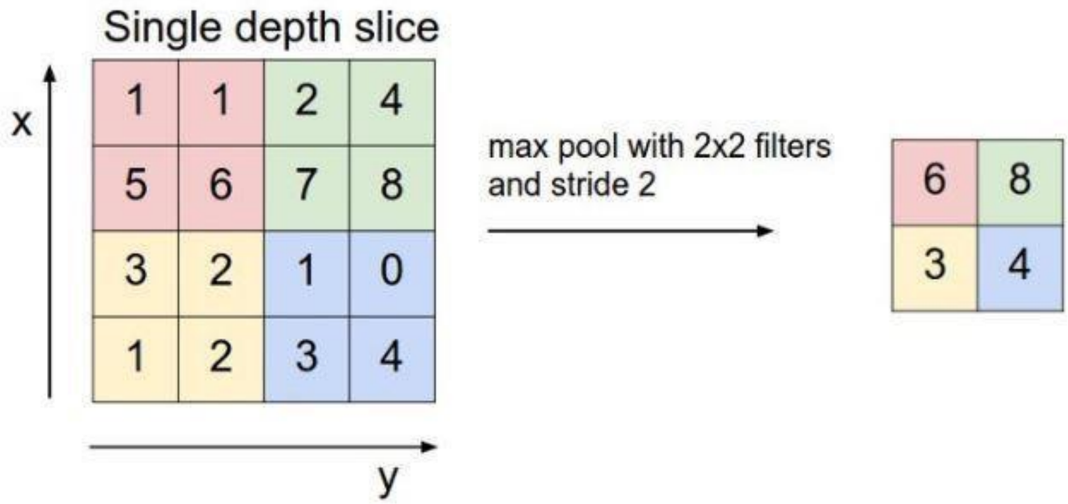


Figure 3.10 Example of the principle of Pooling

Fully Connected Layer:

The fully-connected layer is always the last layer of a neural network. This type of layer receives an input vector and produces a new output vector. To do this, it applies a linear combination and optionally an activation function to the values received as input. The last fully-connected layer is used to classify the input image of the network: it returns a vector of size N, where N is the number of classes in our image classification problem. Each element of the vector indicates the probability for the input image to belong to a class. For example, if the problem is to distinguish cats from dogs, the final vector will be size 2: the first element (respectively, the second) gives the probability of belonging to the class "cat" (respectively "dog"). Thus, the vector [0.9 0.1] means that the image has a 90% chance of representing a cat.

To calculate the probabilities, the fully-connected layer multiplies each input element by a weight, makes the sum, then applies an activation function (logistics if N=2, softmax if N>2): This processing amounts to multiplying the input vector by the matrix containing the weights. The fact that each input value is connected with all output values explains the term fully connected.

$$f^{(l)}[n] = \mathrm{act}\left[b^{(l)}[n] + \sum_{c=1}^{C^{(l)}} < \phi^{(l)}[c], w^{(l)}[n,c] >\right]$$

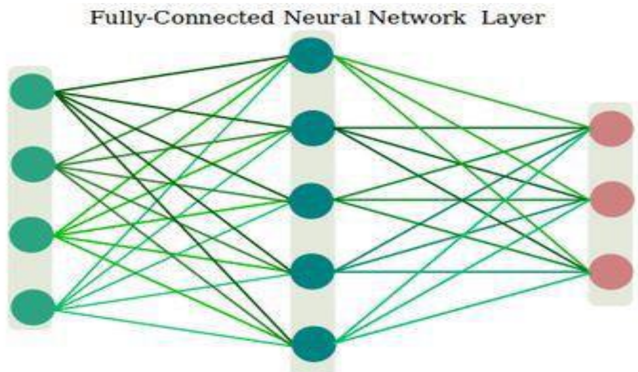


Figure 3.11 Principle of the fully connected layer (fc)

Logistics Layer or Softmax:

Softmax or logistics layer is the last layer of CNN. She resides at the end of the FC layer. Logistics is used for binary classification and Softmax is for multi-classification.

The function is defined as below :

$$\sigma(\mathbf{z})_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}} \text{ pour tout } j \in \{1, \ldots, K\}.$$

CNN Output Layer:

The output layer contains the label which is in encoded form as the shows Figure 3.12.

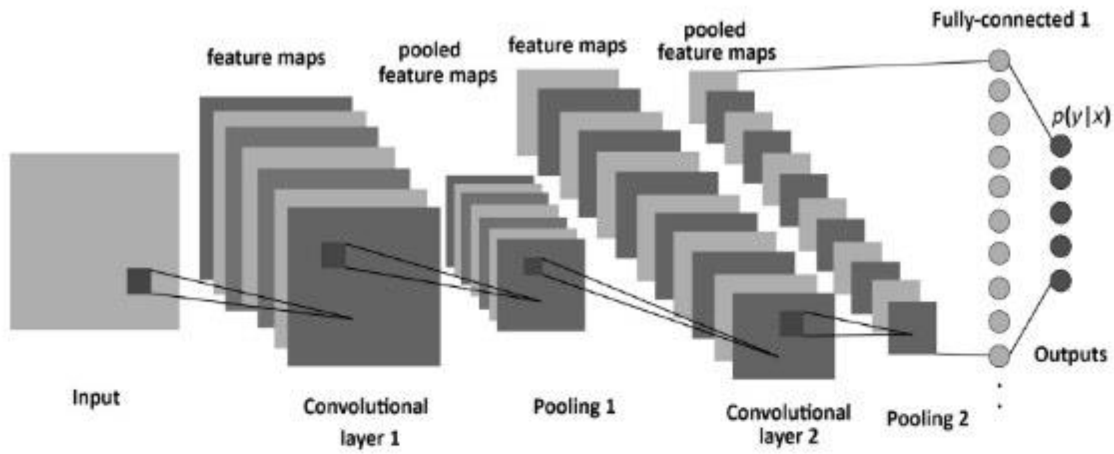


Figure 3.12 Example of the encoded label of the CNN output layer

III.2.3 CNN Architectures:

Till now it has covered the basic concepts of CNN along with different basic components or building blocks of CNN in section 3. Then in section 4, we discuss the learning process of CNN with several learning algorithms with guidelines in order to improve efficiency (including pre-processing, parameter initialization and regularization to CNN). In this section we try to explain some example of successful CNN architecture that shows the recent major advancements in CNN architecture in the computer vision field. Computer vision has three major subdomain where several CNN architectures (models) contribute a vital role to achieve excellent result, we are going to discuss those subdomains with related CNN models as follows. [13]

Image Classification:

In image classification, we assume that the input image contains a single object and then we have to classify the image into one of the pre-selected target classes by using CNN models. Some of the major CNN architectures (models) designed for image classification are briefly described as follows:

LeNet :

The LeNet-5 is one of the earliest CNN architectures, which was designed for classifying the handwritten digits. It was introduced by LeCun et al. in 1998. The LeNet-5 has 5 weighted (trainable) layers, that is, three convolutional layer and two FC layers. Among them, each of first two convolution layer is followed by a max-pooling layer (to sub-sample the

feature maps) and afterward, the last convolution layer is followed by two fully connected layers. The last layer of those fully connected layers is used as the classifier, which can classify 10 digits. The architecture of LeNet-5 is shown in Fig.3.13.

Notes:

- The LeNet-5 was trained on the MNIST digit dataset.
- It used sigmoid non-linearity as the activation function.
- It used stochastic gradient descent (SGD) learning algorithm with 20 training epoch.
- It used 0.02 as the momentum factor value.
- It reduced test error rate to 0.95% on MNIST data set.

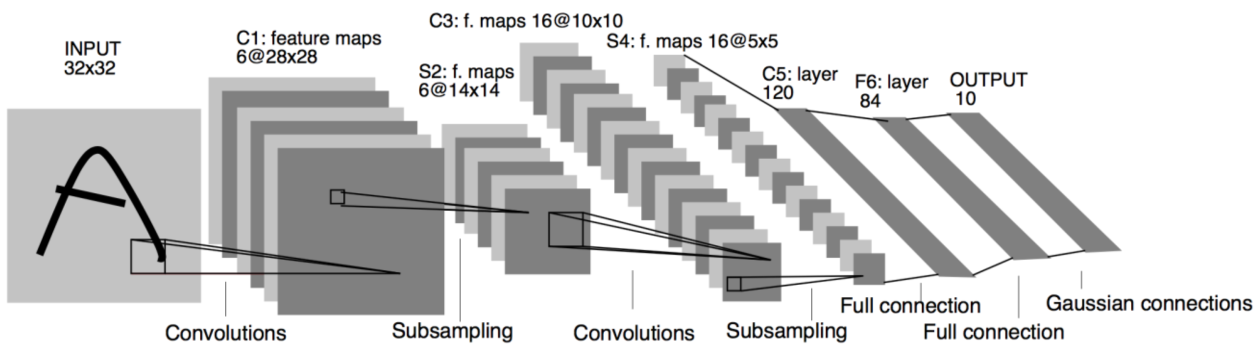


Figure 3.13: The Architecture of LeNet-5.

AlexNet:

Inspired from LeNet, Krizhevky et al. designed first large-scale CNN model, called AlexNet[20] in 2012, which is designed to classify ImageNet data. It consists of eight weighted (learnable) layers among which the first five layers are convolutional layers, and afterward, the last three layers are fully connected layers. Since it was designed for ImageNet data, so the last output layer classify the input images into one of the thousand classes of the ImageNet dataset with the help of 1,000 units. The architecture of AlexNet is shown in Fig.3.14.[14]

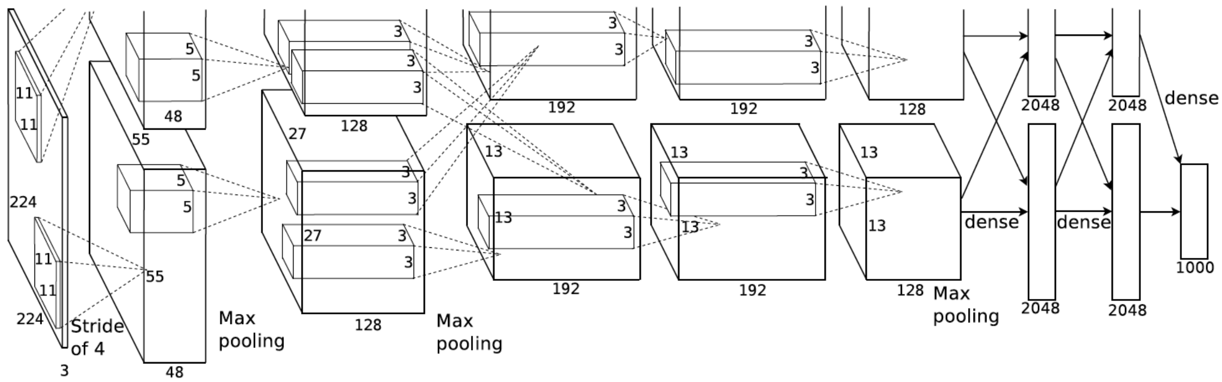


Figure 3.14: The Architecture of AlexNet.

Notes:

• The AlexNet used rectified linear unit (ReLU) non-linearity activation function after each convolutional and fully connected layer.
• It used max-pooling layer after each LRN layer and the last convolutional layer.
• Since it has a larger number of weights (learnable), so to avoid over-fitting it uses several regularization tricks like dropout and data augmentation.
• The AlexNet was trained using stochastic gradient descent (SGD) learning algorithm with min-batch size 128, weight decay 0.0005 and momentum factor value 0.9.
• The AlexNet was trained (on the ImageNet dataset) in two NVIDIA GTX 580 (with 3 GB memory) using cross-GPU parallelization and it takes around six days to complete.
• AlexNet was the winner of ILSVRC-2012.

GoogleNet:

The GoogleNet architecture is different from all the previously discussed conventional CNN models, It uses network branches instead of using single line sequential architecture. The GoogleNet was proposed by Szegedy et al. in 2014. The GoogleNet has 22 weighted (learnable) layers, it used "Inception Module" as the basic building block of the network. The processing of this module happens in parallel in the network, and each (a simple basic) module consist of 1 × 1, 3 × 3 and 5 × 5 filtered convolution layers in parallel and then it combines their output feature maps, that can resulted in very high-dimensional feature output. To solve this issue they used inception module with dimensionality reduction in their network architecture instead of the naive (basic) version of inception module (as shown in Fig.3.15).

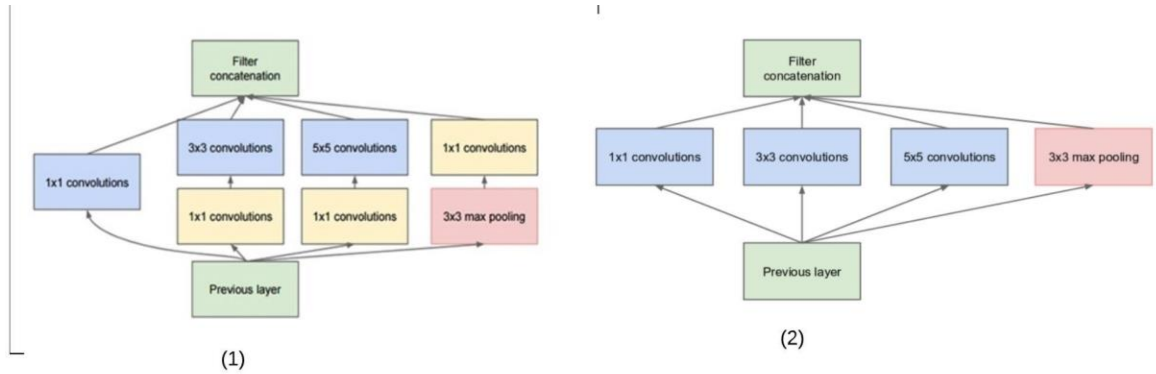


Figure 3.15 Simple Inception Module

Notes:

• Although the GoogLeNet has 22 layers, but it has 12 times lesser parameters than AlexNet.

• It has auxiliary classifiers, that is use to combat vanishing gradient problem.

• It also used rectified linear unit (ReLU) non-linearity activation function.

• It used an average pooling layer instead of the fully connected layers.

• The GoogLeNet used SGD learning algorithm with a fixed learning rate and with 0.9 as momentum factor.

• The GoogLeNet was the winner of ILSVRC-2014

ResNet:

Since a deep CNN model suffers from vanishing gradient problems as we discussed earlier, He et al. from Microsoft, introduced the idea of “identity skip connection” to solve vanishing gradient problem by proposing the ResNet model. The ResNet’s architecture use residual mapping (H(x) = F (x) + x) instead of learning a direct mapping (H(x) = F (x)) and these bolcks are called residual bocks. The complete ResNet architecture is consist of many residual bocks with 3 × 3 convolution layers. Fig.3.16 illustrates the difference between the direct mapping and the residual mapping.

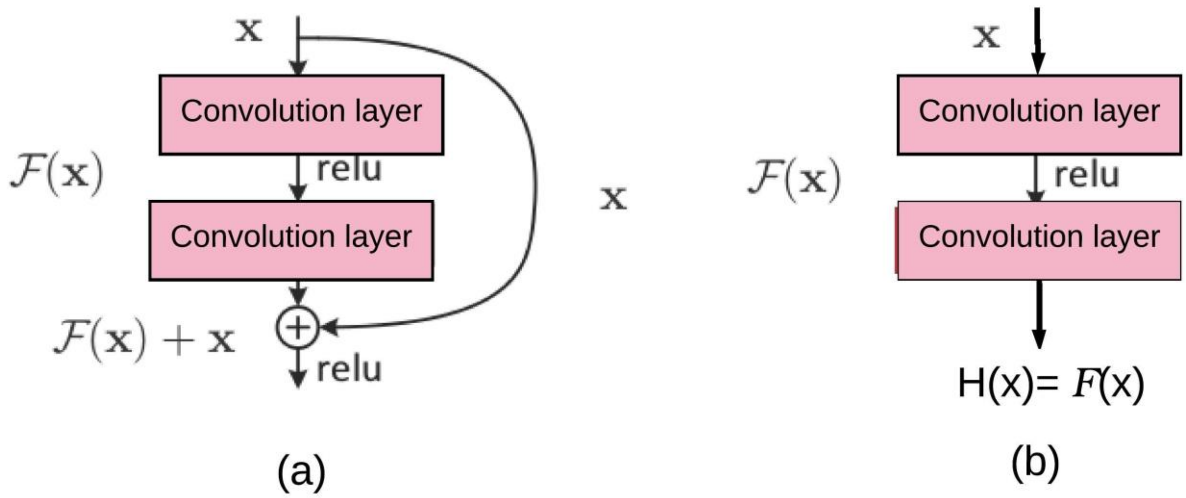


Figure 3.16: (a) Mapping inside Residual block, (b) Simple direct mappings.

Notes:

• The authors propose several version of ResNet with different depth, and they also used 'bottleneck' layer for dimensionality reduction in each ResNet architecture that has depth more than 50.

• Although the ResNet (with 152 Layer) is 8 times deeper than VGGNets (22 layers), it has complexity lower than VGGNets (16/19).

• The ResNet used SGD learning algorithm with the min-batch size of 128, weight decay of 0.0001 and momentum factor of 0.9.

• The ResNet was the winner of ILSVRC-2015 with a big leap in performance, it reduces the top-5 error rate to 3.6% (the previous year's winner GoogleNet has the top-5 error rate to 6.7%).

Conclusion

in this chapter we give some details about deep learning and the method we followed in the FR

we talked the convolutionnal neural network (architecture,layers..) with an explanation of how it works

Chapter 04

Experiments and results

1. Introduction

We have done an implementation of the CNN methodology in order to do an experimental. In this chapter we provide the technical details of the experimental and the results obtained

2. Requirement analysis

Requirement analysis is a process of precisely identifying, defining, and documenting the various requirements that are related to a particular business objective. Requirements gathering helps in clearly understanding the needs of the customer, defining the scope of the project, and resources required to complete it. The functional, non-functional and technical requirements for this project are:

2.1. Functional requirements:

The functional requirement refers to "any requirement which specifies what the system should do". The functional requirements for this project are mentioned below:

- It should be able to handle 'png' and 'jpeg' images.
- It should generate the dataset properly.
- It should be able to predict the authorized users with high accuracy.

2.2. Non-functional requirements:

The non-functional requirement refers to "any requirement that specifies how the system performs a certain function. They are the characteristics or attributes of the system that can judge its operation. The non-functional requirements for this project are mentioned below:

- The GUI of the system will be user friendly.
- The system will be flexible to changes, e.g. an authorized user can be added at any time.
- Efficiency and effectiveness of the system will be made sure.

2.3. Technical requirements:

The technical requirements for this project are mentioned below:

2.3.1. Hardware Requirements:

- CPU : intel i3
- GPU : AMD series 820

- RAM : 8GB
- Hard disc : 128 SSD

2.3.2. Software Requirements:

- Windows operating system
- MATLAB 18 or higher
- Resnet 18 module

3. Generalities on MATLAB

3.1 General definition :

MATLAB is both a calculation software and a high-level programming language. This is paid software, of which there are two free equivalents - Octave is software that uses the language of MATLAB and can therefore use functions written in MATLAB. It is slower and a little less beautiful. - Scilab is developed by INRIA and the syntax differs a little from that of MATLAB but the spirit is the same. It is from my point of view still a little less practical than MATLAB. MATLAB is a numerical calculation software not of algebra unlike Maple. He can only solve numerical equations. The name MATLAB comes from Matrix Laboratory. In MATLAB, all objects are matrices by default. A real variable is therefore seen by MATLAB as a 1×1 matrix. The product is therefore by default a matrix product. You must beware. The type of the variables is not very important. MATLAB can add a Boolean and a real, multiply an integer by a complex without problem. To start MATLAB, just run the MATLAB command in a terminal. This opens the main MATLAB window. We can directly launch command lines there but most of the time we will go through the MATLAB editor which allows you to create scripts [15]

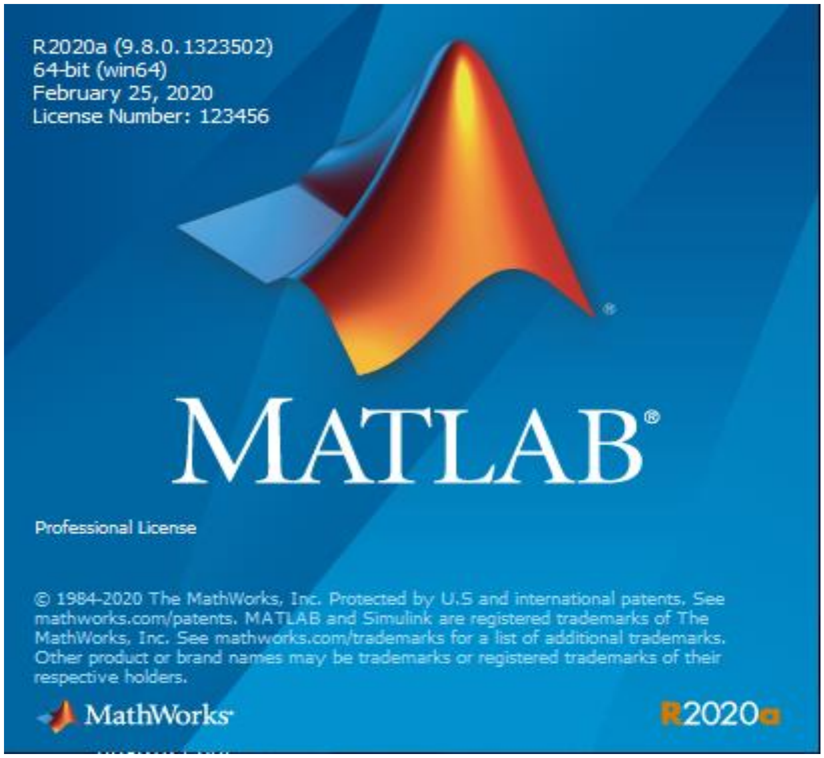


Figure 4.1 MATLAB window

3.2 the particularities of MATLAB

MATLAB allows interactive work either in command mode or in programming mode; while still having the possibility of making graphic visualizations. Considered as one of the best programming languages (C or Fortran), MATLAB has the following particularities compared to these languages:

• easy programming,

• continuity among whole, real and complex values,

• the wide range of numbers and their details,

• the very comprehensive mathematical library,

• the graphical tool which includes graphical interface functions and utilities,

• the possibility of linking with other classic programming languages (C or Fortran).

3.3 Writing a MATLAB program

In MATLAB, programs end with an ".m" extension in the name of the program file. No compilation is to be done before running the program. During execution, an error message appears indicating the locations of the errors. To start the execution of the program, you must always go to the same directory where this program is located. Example: the latter is located in c: \ user; you must first change the directory after launching MATLAB by searching the containing folder the data files are saved with a “. mat" extension and the variables are saved in double precision.

4. Generating dataset:

4.1. Dataset definition:

A machine learning dataset groups together a set of data. These depend on a variable associated with the values. Their access can occur individually or collectively. There are different models, such as the training dataset, the test dataset and the validation dataset.

4.2. Yale faces dataset:

The Yale Face Database (size 6.4MB) contains 165 grayscale images in GIF format of 15 individuals. There are 6 images per subject, one per different facial expression or configuration: center-light, w/glasses, happy, left-light, w/no glasses, normal, right-light, sad, sleepy, surprised, and wink.

The database is publicly available for non-commercial use. In order for us to track those using it.

Figure 4.2: Yale faces dataset

4.3. Dataset resize:

in order to be compatible with Resnet18 the dataset pictures must be in the size of 224x224x3 to do that we run the algorithm below:

```
digitDatasetPath = fullfile('C:\Users\A-Razak\Desktop\AR');
imds = imageDatastore(digitDatasetPath, ...
    'IncludeSubfolders',true,'LabelSource','foldernames');
%outputSize = [224 224 3];
%auimds = augmentedImageDatastore(outputSize,imds);
%img2=zeros(224,224,3);
for i=1:length(imds.Labels)
img=readimage(imds,i);
 %imshow(img2)
img1 = imresize(img,[224 224]);
img2(:,:,1)=img1;
img2(:,:,2)=img1;
img2(:,:,3)=img1;
 imwrite(img2,cell2mat(imds.Files(i)))
end
```

5. Load dataset:

In this section of the code, we load the dataset to imds as below:

```
digitDatasetPath = fullfile('C:\Users\A-Razak\Desktop\CNN2\yalefaces');

imds = imageDatastore(digitDatasetPath, ...
    'IncludeSubfolders',true,'LabelSource','foldernames');
```

6. Split dataset to training and test:

We split the dataset into

- 70% For training
- 30% for testing

```
[imdsTrain,imdsValidation] = splitEachLabel(imds,0.7,'randomized');
```

7. Training process and saving network:

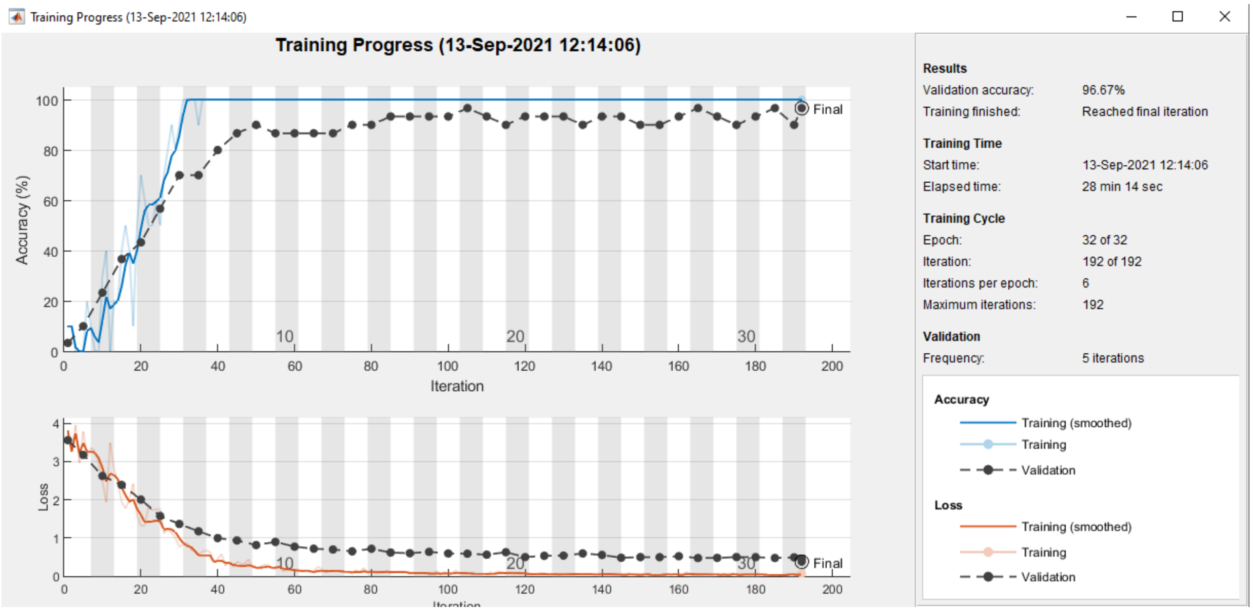
Now we launch the training process to be uploaded to a network which will be saved by ‘save’ command.

```
options = trainingOptions('sgdm', ...
    'MiniBatchSize',10, ...
    'MaxEpochs',20, ...
    'InitialLearnRate',1e-4, ...
    'Shuffle','every-epoch', ...
    'ValidationData',imdsValidation, ...
    'ValidationFrequency',5, ...
    'Verbose',false, ...
    'Plots','training-progress');
trainedNet = trainNetwork(imdsTrain,lgraph,options);

[YPred,probs] = classify(trainedNet,imdsValidation);
YValidation = imdsValidation.Labels;
accuracy = sum(YPred == YValidation)/numel(YValidation)
save trainedNet ;
```

We can change the epochs numbers in order to get higher accuracy results

The result of this process will be 96.67 % as it shown below:



SVM classifier

Support Vector Machine Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. A hyperplane is called to the best decision boundary. SVM chooses the extreme points that help in creating the hyperplane. These extreme cases are called as support vectors, and the algorithm is termed as Support Vector Machine. Consider the below diagram in which there are two different cases that are classified using a decision boundary or hyper plane

```
 layer = 'pool5';
featuresTrain = activations(trainedNet,imdsTrain,layer,'OutputAs','rows','ExecutionEnvironment','cpu');
featuresTest = activations(trainedNet,imdsTest,layer,'OutputAs','rows','ExecutionEnvironment','cpu');
whos featuresTrain
 YTrain = imdsTrain.Labels;
YTest = imdsTest.Labels;
classifier = fitcecoc(featuresTrain,YTrain);
YPred = predict(classifier,featuresTest);

accuracy = mean(YPred == YTest)
```

The accuracy result will be the same 96 % as it shown in the capture below:

```
accuracy =

    0.9667
```

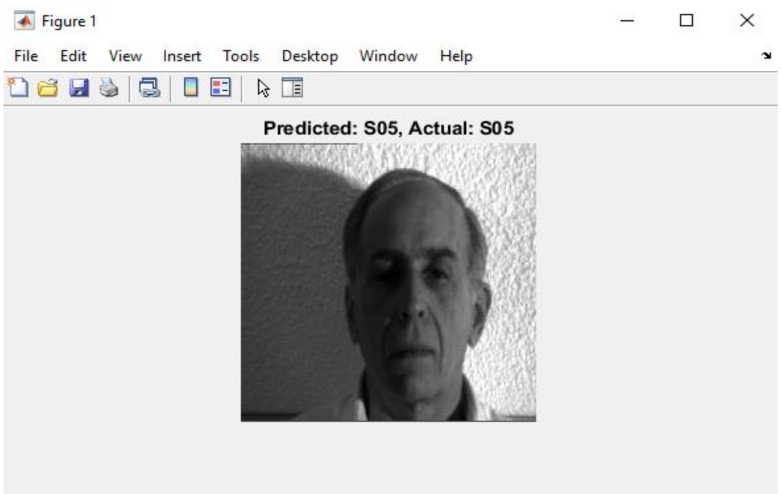
The saved network will be used later in the testing process.

8. Testing process

In this part we use one of the pictures in the dataset and launch the algorithm to see the closest match.

```
img = readimage(imds,10)
actualLabel = imds.Labels(10);
predictedLabel = trainedNet.classify(imgl);
imshow(imgl);
title(['Predicted: ' char(predictedLabel) ', Actual: ' char(actualLabel)])
```

The result will be a matching between given picture and the trained dataset as below:



discussion of results:

as you can see at the epoch 20, we see that the accuracy using CNN classifier is higher the using SVM classifier

so, CNN classifier is the better then SVM when we apply it in yale faces dataset

Conclusion

finally using MATLAB we experiment the CNN method in face recognition using the resnet 18 training function with CNN and SVM classifier with yale faces dataset

General conclusion

We presented in this project all the necessary steps for the development and facilitates the management of attendance via facial recognition technology.

Given the amount of potential software (security, social networks, etc.) that can be based on this application, it must meet the requirements of speed and robustness of the results

In this sense, the first part consists of locating faces uses CNN layers. The second part of the deals with the recognition of localized faces by applying the CNN method.
Convolutional neural networks (CNN) was first proposed in the 1960s, when Hubel and Wiesel discovered its unique network structure, which can effectively reduce the complexity of feedback neural network, while studying the neurons used for local sensitivity and direction selection in the cat cerebral cortex.

Indeed, capturing an image of a face, is simple and non-invasive. It is therefore a biometric modality easily tolerated by users, but the performance of facial recognition is still far beyond what one would expect for such applications.

This project allowed us to discover more deeply several aspects of the development of complex software. We first had to learn about the algorithmic side of face recognition, and more generally of computer vision. which is a growing field. This research therefore led us to the creation of a “raw” facial recognition engine. We had to solve several algorithmic problems having more or less links with mathematics, an important discipline in image processing in general.

Webography:

https://www.researchgate.net/

https://github.com/

https://www.mathworks.com/

https://www.intechopen.com/

References

[1] Anil K. Jain, Ruud Bolle, Sharath Pankanti I. Biometrics: Personal Identification in Networked Society.
[2]R. Jafri, H. R. Arabnia, “A Survey of Face Recognition Techniques”, Journal of Information Processing Systems, Vol.5, No.2, June 2009.
[3]AMEUR, B., S. MASMOUDI, A. G. DER- BEL and A. BEN HAMIDA. Fusing Gabor and LBP feature sets for KNN and SRC-based face recognition. In: International Conference on Ad-vanced Technologies for Signal and Image Process- ing (ATSIP)
[4]Sujata G. Bhele and V. H. Mankar, A Review Paper on Face Recognition Technique, 2012.
[5]MEENA, D. and R. SHARAN. An approach to face detection and recognition. In: Inter-national Conference on Recent Advances and Innovations in Engineering (ICRAIE)
[6]Hua Yu, Jie Yang, A direct LDA algorithm for high-dimensional data with application to face recognition.
[7}S. Asadi, Dr. D. V. Subba R. V. Saikrishna, "A Comparative study of Face Recognition with PCA and Cross-Correlation Technique",
[8]R.S. Michalski, J.G. Carbonell, T.M. Mitchell, Machine Learning: An Artificial Intelligence Approach.
[9]Schmidhuber, J. (2015). "Deep Learning in Neural Networks: An Overview". Neural Networks.
[10]A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks, 2012.
[11]G. MERCIER and R. FABLET. Convolutional 2016, pp. 95–99. ISBN 978-1-5090-2889-4. Neural Networks for object recognition
[12]https://towardsdatascience.com/review-of-lenet-5-how-to-design-the-architecture-of-cnn-8ee92ff760ac
[13] BRITZ, D. Understanding convolutional neural networks. In: WILDML
[14] International Workshop on based on convolutional neural network and sup- Biometrics and Forensics
[15]matlab.org