IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 26, Issue 2, Ser. 2 (Mar. – Apr. 2024), PP 58-70
www.iosrjournals.org

Facial Recognition System Based On The Convolutional Neural Network And Opencv For Automatic Attendance At The Civil Service Of Bandundu Provincial Government Dr Congo

Jean Chrysostome Mayoko[1,3], Aser Itakala [1], Chadrack Lubamba[1,3], Sllife Nyazabe[1,4], Espérant Kikata[3], Hervé Lamanabwe[3], Serge Tamina[2]

[1]department Of Business English And Computer Science (Igaf), Faculty Of Economics Science And Management, University Of Kinshasa, Kinshasa, Dr Congo
[2]department Of Computer Science, Graduate Business School, Bandundu, Dr Congo
[3]department Of Mathematics And Computer Science, Faculty Of Science And Technology, University Of Kikwit-Drc
[4]graduate School Of Global Digital Innovation (Gdi), Korea Advanced Institute Of Science And Technology (Kaist), Daejeon, South Korea

Abstract
This study aims to propose a safer and more secure attendance monitoring system for Civil servants in the Democratic Republic of the Congo based on computer vision applying the OpenCV library and deep learning using the convolutional neural network related to facial recognition and identification. The convolutional neural network model's training results indicate 100% accuracy in recognition while the evaluation model yields 100% ability in prediction on the images used. Furthermore, the integration of the obtained model in the OpenCV library, accurate outcomes and was effectively used to recognize the staff members based on the images laid out from the camera by distinguishing them with a low margin of error dependent on the quality of the light.
Based on the positive results obtained during this study, we can conclude that integrating this kind of system into the civil service of the DRC's provincial government will positively increase agent performance and productivity as an effective system of attendance check compared to the mechanisms currently in use, which are covered with numerous irregularities.

Keywords: *Deep Learning, Computer Vision, Image Segmentation and Detection, OpenCV, Convolutional Neural Network*

Date Of Submission: 26-03-2024 Date Of Acceptance: 06-04-2024

I. Introduction

The performance of agents in a firm is one of the most important factors in allowing the organization to remain competitive in a market where competition is increasing. The presence of agents on the workstation and on time is a key aspect of performance measurement because we live in an information-driven environment. It is critical to compel employees to arrive on time and not be absent without reason using dependable and secure techniques, as this has a favorable impact on increasing service quality and, most importantly, performance.

Computer science, the discipline of automatically and rationally processing information, plays an important role in modern human existence. It provides a wide range of solutions, from simple enjoyment (smartphone games, sound and video players, etc.) to complex applications.

Man lives in an era in which computing is at his service where enormous quantities of information are processed continuously at every moment and transported around the world via networks interconnected by the Internet.

Artificial intelligence (AI) refers to machines that mimic the cognitive functions of human intelligence. AI applications are being used in a variety of fields, including object detection, picture categorization, speech recognition, and translation.

Machine learning, a subset of artificial intelligence, employs algorithms taught using mathematical and statistical methodologies to execute tasks by learning patterns from data rather than obtaining these skills through coding (Chartrand G., 2017). Biological neural networks spurred the development of artificial neural networks, a subfield of machine learning. Various artificial neurons are linked to build a network of tiered processing units.

Between the input and output layers, there are a number (at least one) of hidden layers that are in charge of network decision-making (Chartrand G., 2017) (Erickson BJ., 2017).

The artificial neural network (ANN) was inspired in general by the human nervous system, which is composed of neurons that process information. A neuron gets information from another neuron through its dendrites and transmits it to other neurons via its axon. According to (Bishop M., 1995), the biological brain contains around 10 billion neurons, with each neuron connecting to others via 10,000 synapses. ANN guarantees effective learning using its many ways. The presence of differential equations (gradients), random selection, and weight updates are critical to the success of ANN.

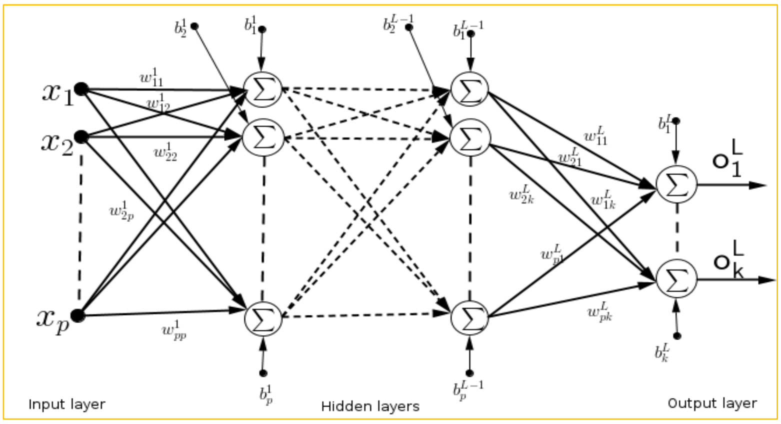


Figure 1: Architecture of an ANN (Masakuna J.F., 2019)

A network has exactly one input layer and one output layer, and the number of hidden layers depends on the recognition model problem; nevertheless, using a hidden layer according to (Nielsen M., 2023).

The ANN architecture, shown in Figure 1 above, seeks to identify a set of weights w that will precisely correspond to the desired output. As a result, several mechanisms can be applied.

Multilayer Perceptron (Multilayer Perceptron): With this method, each neuron, except those of the input layer, performs the sum of the information received and activates the sum, that is to say, it puts the sum scaled to a real value in [0; 1]. The model in Figure 1 as defined, designates for each neuron the same activation function $g(.)$and a bias of its own, therefore helping to move the output from 0 to 1 or vice versa.

In his research, (Nielsen M., 2023) recommends using MLP when nonlinearly separable data is processed. The MLP architecture contains at least one hidden layer, as shown in Figure 1.

The cost function: the cost function is a measure of the error between the value predicted by the model and the actual value.

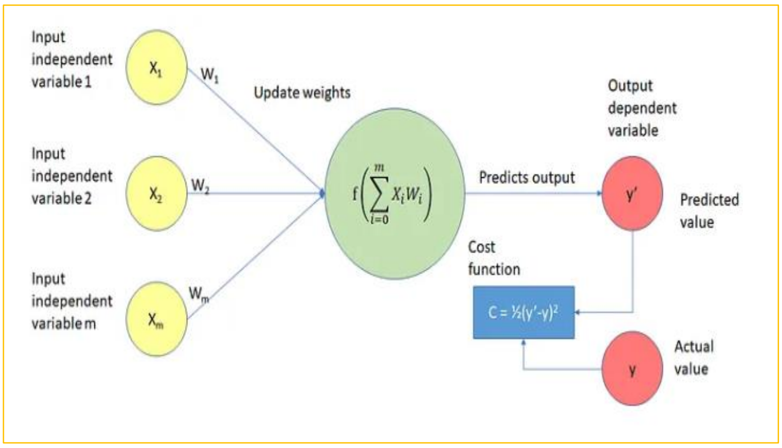


Figure 2: ANN with details on the cost function (Mohammed Zeeshan M., 2020)

For example, we want to predict the value y_i for the data point x_i.

$f\theta(x_i)$ (1)

Represent the prediction or output of an arbitrary model for the point x_i with parameters θ. One of the many cost functions can be given by:

$\sum_{i=1}^{n}(y_i - f\theta)^2$ (2)

The search (Nielsen M., 2023) thus defines the cross entropy function by:

$$E_c(w,b) = -\frac{1}{n}\sum_i \sum_j \left[y_j(x_i)\ln o_l^L(x_i) + (1-y_j(x_i))\ln\left(1-o_l^L(x_i)\right)\right], \quad (3)$$

Or $o_j(x_i) = g(z_j) = g(\sum_k w_{jk}x_j + b_j)$ and $g(.)$ the sigmoid activation function given by $g(z_j) = [1 + \mathrm{Exp}(-z)]^{-1}$. It follows with:

$$g'(z) = g(z)[1-g(z)] \quad (4)$$

Entropy has the advantage that the closer it gets to zero as the neuron's output approaches the exact output and the larger the error, the faster the classifier (unlike the quadratic cost function for example) (Masakuna J.F., 2019). The changes in $w$and $b$are determined using the backpropagation (BP) approach which introduces the intermediate quantity δ_i^l representing the error in the $i^{ème}$layer neuron l.

backpropagation technique or Backpropagation (BP) in English gives the procedure to calculate δ_i^l. Consider an input $z_j^l$and output o_j^l given by:

$$z_j^l = \sum_k w_{jk}o_k^{l-1} + b_j^l \text{ et } o_j^l = g(z_j^l) \quad (5)$$

The classification approach guided by multilayer neural networks (perceptron) learns from a set of prototype data (class instances), each of which is described by a set of attributes. The multilayer perceptron's performance is limited because of the difficulty in selecting attributes and the often modest number of layers (1 to 2 layers) (Rumelhart R. J. and Hinton G. E., 1986). One of the most common forms of deep neural networks is the Convolutional Neural Network (CNN), which uses convolutional layers to extract characteristics from data. Its architecture is perfect for processing raster data such as photographs, and CNN eliminates the need for manual feature extraction, which is typically a time-consuming procedure. Based on the multilayer perceptron (MLP) and inspired by the behavior of vertebrates' visual cortex, convolutional neural networks successfully adapt to the main faults of MLPs and are built of multiple layers, as shown in Figure 3.

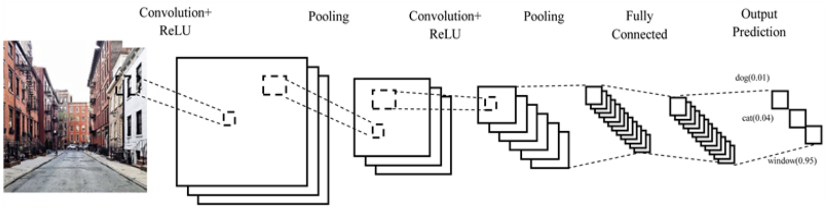


Figure 3: Architecture of a CNN (Ren J. and Wang Y., 2022)

Convolutional networks have a significant advantage in that they employ a single weight for all signals entering the same convolution kernel. This approach decreases memory footprint, increases performance, and allows for translational processing invariance. This is the primary advantage of a convolutional neural network over a multilayer perceptron, which considers each neuron independent and hence assigns a separate weight to each incoming signal.

Let us therefore point out that neural networks and deep learning currently offer the best solutions to many problems in image recognition, speech recognition, and natural language processing (Nielsen M., 2023) (SAIDI R., 2023). The result of a study conducted by (Ren J. and Wang Y., 2022)on several object detection algorithms, reveals that deep learning technology has a great success and made breakthroughs in science Besides, there is still a huge gap between the effectiveness and the speed of the detection model as well as humanized performance. According to research (Krizhevsky, 2017), convolutional neural networks (CNN) have made significant progress in recent years and are a bright spot in the developing treasure trove of deep neural networks. In addition, a detailed investigation of deepfake production and detection technologies was undertaken, with deep learning algorithms used to successfully detect fake images and videos. It concludes that deep learning approaches are exceptionally effective at detecting deepfakes. Furthermore, the number of images required for a better outcome remains a constraint, as does the challenge of determining the number of layers required and which architecture is most suited for detection (Abdulqader M., 2021).

Even though, when it comes to spatial interpolation of 3D structures, the convolutional neural network model outperforms other methods. The research concludes (RAKOTONIRINA H., 2023). The concepts of artificial intelligence and computer vision are closely related. Humans use vision to adapt to and understand their surroundings, whereas computer vision duplicates human vision but perceives and interprets images electronically.

Computer vision not only serves as an eye to see, but it also must respond. It must be capable of detecting, identifying, and processing seen pictures in the same way that human vision does (Teoh K., 2021). Convolutional neural networks (CNNs) perform well in computer vision tasks (Chartrand G., 2017). Computer vision must do three actions known as major, which are similar to what a motorist would do if he unexpectedly encountered a human in his way. The driver must react swiftly and act, and the brain passes through three stages: detection, processing, and decision-making. Vision, as a fundamental component of intelligence, involves a wide range of skills such as coordination, memory, recall, reasoning, estimating, and recognition. A system with only one of these skills does not meet the criteria for a vision.

Computer vision technology has emerged as a critical area of focus in Internet applications (Ren J. and Wang Y., 2022). As one of the fundamental problems of computer vision, object detection has become the basis of many tasks in this field, and at the moment, object detection is not only the subject of much academic research but is also used in real life, such as video fire detection, according to (Kim B. and Lee J., 2019) and Object identification algorithms are classified into two types: classic object detection methods based on image processing and convolutional neural networks. Note that the convolutional neural network (CNN) has made significant development in recent years regarding the detection problem utilizing images as input data and has been widely used in computer vision (Simonyan K., 2014) (Long J., 2015) (KHOUJA, 2015).

Furthermore, according to (Teoh K., 2021), the human face is a significant attribute for identifying a person because everyone has a unique face, including twins. Thus, facial recognition and identification are required to differentiate oneself. Facial recognition is now widely used in a variety of applications, including phone unlocking, criminal identification, and even home security systems. Similarly, image processing and feature extraction are critical steps in the object detection process. Digital image processing encompasses all approaches for changing digital images to improve or extract information from them. Image processing includes mathematical approaches for preprocessing an image to eliminate noise. And feature extraction uses mathematical approaches to alter an image in order to get essential features for accurate categorization.

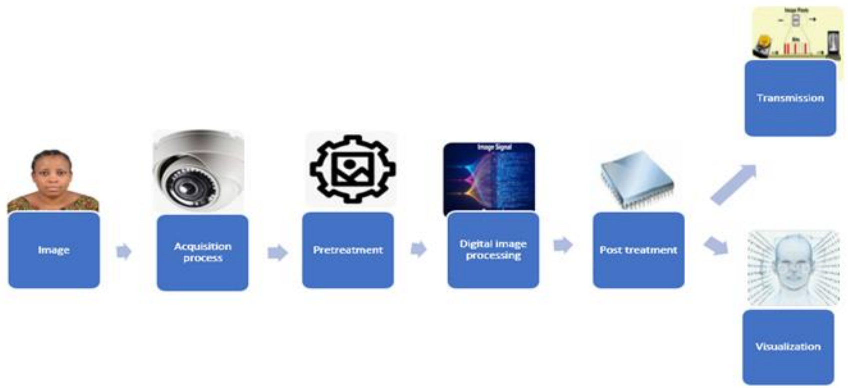


Figure 4: Diagram of an image processing system

Consider the raw data X and query imagine x_q as input to obtain the output X^c and x_q^c. The image can thus be transformed using a variety of mathematical techniques.

Dimensionality and filtering

The first stage in image processing considered in this study is filtering, which involves lowering the size of the image before proceeding with any other process by changing its mode. In the default image mode, each pixel in an image has five dimensions (x, y, i, j, k), where (x, y) is the pixel's position and (i, j, k) are the color intensities of the code function for Red, Green, and Blue (RGB). The new mode (grayscale mode) to define has three dimensions (x, y, f), where f represents the intensity function. A base (x, y, i, j, k) to base-mapping (x, y, f) with the function $f(x, y)$defined as follows:

$$f(x, y) = \frac{\alpha i(x,y) + \beta j(x,y) + \gamma k(x,y)}{\mu} \quad (6)$$

Or $\alpha, \beta, \gamma \in [0,255]$ and $i, j, k \in [0, 255]$. $\mu \in [1, 255]$. The coefficients $\alpha, \beta, \gamma$are defined as follows $\alpha + \beta + \gamma = 1$. So,$f(x, y) \in \begin{cases} [0, 1], & \mu = 255 \\ [0, 255], & \mu = 1 \end{cases}$

Thus, all colors can be obtained from the basis (i, j, k). After dimensionality is reduced, filtering is employed to eliminate noise from the image, such as with the median filter. According to (Jassim F. A. and Altaani F. H., 2013), in the median filter, each pixel is replaced by the median $g(x, y)$ of its nearest neighboring pixels; in other words, a representative pixel $g(x, y)$ replaces each pixel. The user determines the number of neighbors to consider, and it is preferable to select an odd number such that the median is exactly one of the current pixels. After this filtering, all of a pixel's nearest neighbors have the same grayscale intensity function $g(x, y)$.

Binarization

Noise reduction is the initial step in image processing, followed by segmentation, which is the process of identifying various shapes in an image. Based on the segmentation technique's requirements, we obtained two separate sets G (foreground set) and B (background set) based on the gray level intensities of the pixels $g(x, y)$. To derive B and G, some researchers (Jassim F. A. and Altaani F. H., 2013) compare each pixel to the average pixel θ, considered as the threshold. Binarization can be mathematically expressed as follows:

$$h(x, y) = \begin{cases} 0, & si\ g(x, y) \leq \theta, ex.\ g(x, y) \in \mathcal{B} \\ 1, & sinon, ex.\ g(x, y) \in \mathcal{G} \end{cases} \quad (7)$$

Where $h(x, y)$ is a binarized function. Unfortunately, this method does not take into consideration image variations. Therefore, Otsu proposes an ideal strategy to find the threshold that takes into consideration pixel variance (Jassim F. A. and Altaani F. H., 2013). In computer vision and image processing, the Otsu method is used to mechanically perform image thresholding based on the shape of a histogram, or the scaling down of a gray level image to a binary image (Makkar H. et Pundir A., 2014). This method searches for the threshold which minimizes the intra-class variance σ_{in}^2 (t)and maximizes the interclass variance $\sigma_{out}^2(t)$in the image where $t$any threshold is. If $t_{min} = \min(x_q)$, $t_{max} = \max(x_q)$, $\omega_b(t) = \sum_{p=t_{min}}^{t-1} \pi(p)\ et\ \omega_a(t) = \sum_{p=t}^{t_{max}} \pi(p)$, thus,

$$\sigma_{in}^2(t) = \omega_b(t)\sigma_b^2(t) + \omega_a(t)\sigma_a^2(t) \text{And} \sigma_{out}^2(t) = \omega_b(t)[\mu_b(t) - \mu]^2 + \omega_a(t)[\mu_a(t) - \mu]^2 \quad (8)$$

Where $\sigma_b^2(t)$ and $\sigma_a^2(t)$ are variances of the pixels in $\mathcal{B}\ et\ \mathcal{G}$respectively, $\mu_b(t)$ and $\mu_a(t)$ are respectively the mean of the pixels in $\mathcal{B}\ et\ \mathcal{G}$, μ is the mean of the pixels in the image $\pi(p)$the probability of each pixel $p$in the image x_q.

Likewise, this method can therefore be explained thus, as mentioned above, we look for the threshold which reduces the intra-class variance (within the class variance), classified as the weighted sum of the variances of the two classes:

$$\sigma_\omega^2 = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t) \quad (9)$$

Where the weights ω_i are the different probabilities of two classes partitioned using the threshold $t$and σ_i^2 are the divergences of these sets.

The idea behind proposed by Otsu is that by seeking to minimize intra-class variances, in other words, it amounts to maximizing the inter-class variance as follows (Makkar H. et Pundir A., 2014):

$$\sigma_b^2(t) = \sigma^2 - \sigma_\omega^2(t) = \omega_1(t)\omega_2(t)[\mu_1(t) - \mu_2(t)]^2 \quad (10)$$

Which is expressed in terms of class $w_i$and middle class probabilities μ_i. And the class probability $\omega_1(t)$ is therefore calculated from the histogram t:

$$\omega_1(t) = \sum_0^t P\ i \quad (11)$$

Where the class average $\mu_1(t)$ is given by:

$$\mu_1(t) = \sum_0^t P(i)x\ i \quad (12)$$

Where, x (i): refers to the value at the center of i in the histogram bin. Likewise, $\omega_2(t)$ and μ_t on the right side of the histogram, the bin is larger than what can be calculated. Class probabilities and averages can be determined iteratively.

Otsu algorithm

The OA (Otsu Algorithm) is then given by (1) Calculate the probabilities $\pi(p)\ \ \forall p \in x_q$. (2) Calculate $\sigma_{in}^2(t)$ and $\sigma_{out}^2(t)$ for all thresholds t. The optimal threshold θcorresponds to $t$which has the maximum of $\sigma_{out}^2(t)$and the minimum of $\sigma_{in}^2(t)$. If there are several optimal thresholds, it is rather their average which is taken into account.

This research includes the enhanced Otsu algorithm, which is based on a median threshold proposed by (Yang X., 2012) (Hima Bindu Ch. et Satya Prasad K., 2012) (Zhan Y. et Zhang G., 2019) to overcome the result quality problem when the histogram of a gray level is not bimodal or noise reduction.

Segmentation

After binarizing the image by detecting both sets $\mathcal{B}\ et\ \mathcal{G}$using OA as described above, the next step is image segmentation x_q, i.e. searching for different shapes in x_q. In image segmentation, each pixel$pi \in x_q$ is assigned exactly to one form $x_q^j$of x_q. i.e. find the different forms $x$in $\mathcal{G}$according to the probability γ_i^j(a probability that pibelongs in x_q^j). The algorithm that will then be used is the one based on graph theory (Grady L., 2006)and calculates the probability of each graph vertex belonging to the different shapes. The image $x_q$is therefore represented by the graph $G = (P, H)$ where $pi \in P$ defines as a set of pixels and $h_{ij} = (p_i, p_j) \in H$ defines as a set of edges.

The function of randomly mapping an image x_q into a graph $\mathcal{G}$ of the segmentation technique based on the Random Walker Algorithm (RWA) proposed by (Grady L., 2006), motivated the choice of this algorithm.

The random walk algorithm takes an image as input, and forms the undirected graph G = (N, E) (Feng A., 2023). Nodes N are the voxels of the image and edges E designate the connection between two neighboring nodes. This algorithm formulates the following two hypotheses: (1) G is a connected and undirected graph. (2) The neighbors of a pixel are chosen randomly. Note that the Graph can be weighted and the segment x_q^j of the image x_q can thus be found:

The weights of the graph G are defined by the Gaussian function n_{ij}

$$n_{ij} = \exp\left(-\frac{[g(x_i,y_i)-g(x_j,y_j)]^2}{\sigma}\right) \quad (13)$$

Where $g(x_i, y_i)$ is the image intensity at the node p_i and σ is the standard deviation in x_q and from (Grady L., 2006), RWA optimizes the following energy:

$$Q(\gamma) = \Upsilon^T M \Upsilon \quad (14)$$

Where the semi-positive Laplacian matrix M formed by the graph G is given by:

$$M = (\mu_{i,j})_{n_x n}, \quad \mu_{i,j} = \begin{cases} d(p_i), & si\ i = j \\ -n_{ij}, & si\ \neq j\ et\ p_i\ esst\ adjacent\ à\ p_j \\ 0, & sinon \end{cases} \quad (15)$$

Where γ_i is the sequence of vertex probabilities p_i with respect to all shapes x_q^j found in x_q and $d(p_i)$ is the vertex degree p_i. And the probability matrix Υ is given by $\Upsilon = [\gamma_1,\ \gamma_2, \dots ,\ \gamma_n]^T$.

Thus, The RWA is given as follows: [1] Apply binarization to obtain the sets ($\mathcal{B}\ et\ \mathcal{G}$). [2] Map the graph (image) using equation (13) above. [3] Solve equation (14). [4] The vertex p_i is assigned to the shape x_q^k such that $\gamma_q^k = max\{\gamma_i^j\}\ j \geq 1$. To note that $\sum_j \gamma_i^j = 1$

II. Methodology

This research aims to establish a facial recognition system for civil service agents in the Democratic Republic of the Congo. To create such a system, an examination of the existing system was made, afterwards trained a facial recognition model based on CNN, and lastly used the learned model to detect, identify, and report the attendance of each staff member.

Analysis of the existing system

The civil service is a big corporation in the Democratic Republic of the Congo that hires practically all state officials assigned to other public services in the country as a whole and the province of Kwilu specifically. Because they handle nearly all of the country's agents, the employees there must display remarkable expertise and a willingness to work. Furthermore, we see evidence of inattentiveness or carelessness on the part of agents who decide to come at the time that suits them and at any moment and go back much earlier than expected because the existing system encourages such behavior.

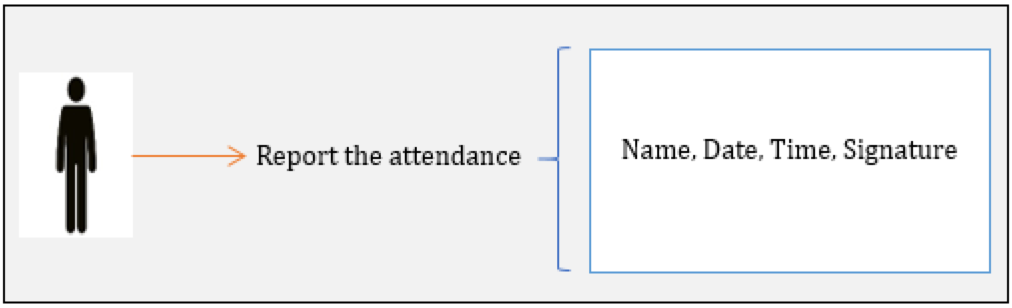


Figure 5: Attendance report in existing system

The most serious issue with the current method of reporting attendance is that it encourages agent performance decline because the majority of agents fail to attend the service and have the register signed by colleagues at work. Others arrived late and left early or before the allotted hour.

Proposed system

Throughout this study, authors suggest an automatic attendance system using Artificial Intelligence, specially a Convolutional Neural Network and OpenCV Library.

Library selection

In general, a human recognition system consists of two phases: facial detection and identification. Object detection is an important computer vision problem that entails recognizing and localizing objects of interest in pictures or movies. It seeks to build an environment tailored to the needs of a blind person, for example, in order to provide optimal comfort and performance in critical situations. The OpenCV library, a free library originally developed by Intel specialized in real-time image processing, is the state-of-the-art real-time object detection technique since it is substantially faster than other algorithms while yet maintaining acceptable accuracy (Nikitha P. et al., 2023). It is a huge open-source library for image processing, machine learning, and computer vision (Sriratana, 2018).

A speed comparison was performed between OpenCV and Matlab (Matuska S., 2012). Figure 5 depicts the basic approach for image processing, with the essential point being the time consumption in OpenCV and Matlab. It was discovered that OpenCV is substantially faster than Matlab by up to 30 times and possibly up to 100 times for the erosion technique.

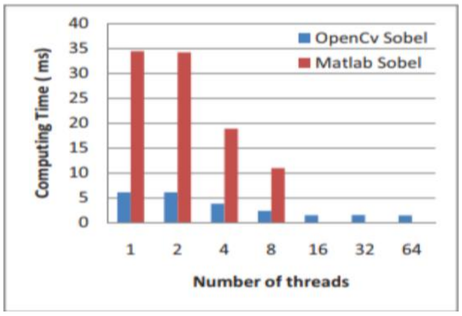


Figure 6: Execution time between OpenCV and Matlab (Matuska S., 2012)

Matlab is written in a high-level language and is built on Java while Java is built on C language. So when a Matlab program is executed, the computer takes time to interpret the coding. It then transforms them into Java code and finally executes the code. On the contrary, OpenCV is generally a library written in C language. The operations performed are major in processing but not in interpretation. So the program in OpenCV can run faster than that in Matlab. In general, OpenCV runs faster than Matlab and is the most comprehensive open source library for computer vision in terms of feature and object recognition, as well as machine learning (Kumar A., 2022). Therefore, OpenCV was chosen to be applied in this paper.

Facial recognition process

After training the model, it was integrated into the OpenCV library to recognize agents and mark their attendance. An image sensor or a basic camera needs to be set up to record or take pictures since facial recognition needs to identify and authenticate a person. The camera needs to be compatible with the program being used.

Creating an input image, which can be a still image, a recorded video, or real-time video, should be the following stage. After providing input, faces in photos or videos are recognized. Once trained, the classifier is utilized to begin the recognition process. It can be used in videos or images to identify one or more persons to mark attendance, one face at a time is taken into account.

Different sets of Python scripts are provided to perform the different recognition types. The Python script will import the classifier trained in the previous step to perform person recognition from the camera or an image.

Once the recognition has been completed, if the face is known by the system, the name, date and time are recorded in a Database folder (Excel file).

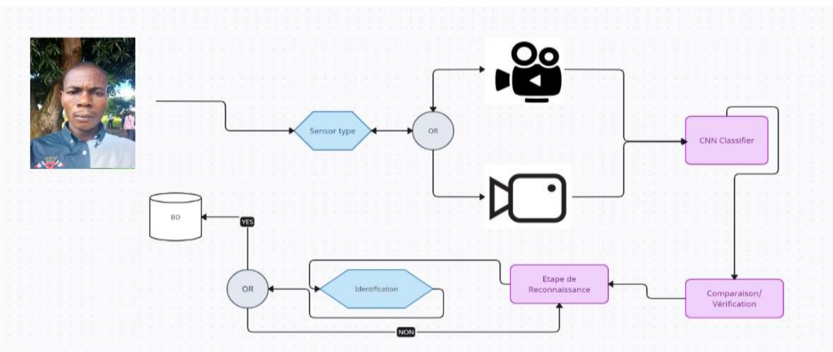


Figure 7: Proposal of the new system

Model training process

The collection of data (images) is the first step of the process, then the preprocessing (separating the training and test data), followed by the creation of the model, finally the validation of the model.

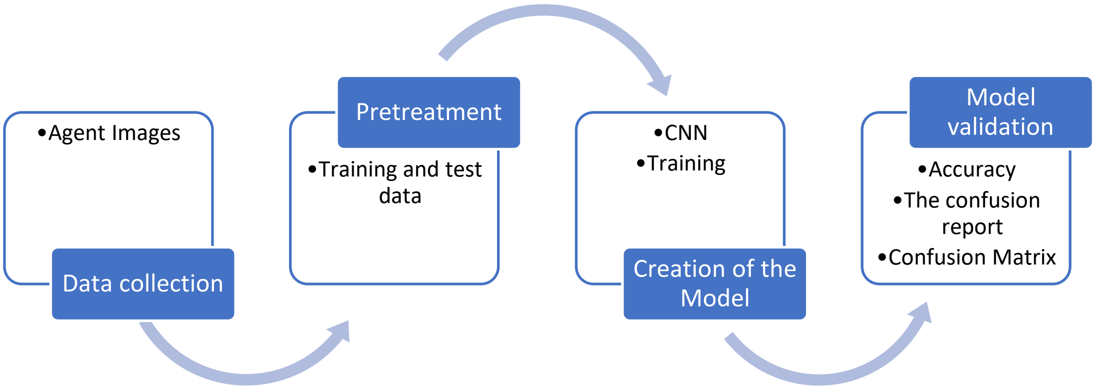


Figure 8: CNN Model implementation process

Classification report is one of the metrics used to evaluate the success of a machine learning based model. It shows the precision, recall, F1 score and support of the algorithm used in this study to better understand the overall performance of the model. The author (KHARWAL A., 2021) (Dimitrovski I., 2023) supports the thought that to understand the categorization report of a machine learning model, it is necessary to become familiar with all of the measures presented in the classification report.

The Confusion Matrix, provides an output matrix and summarizes the overall performance of the model. The confusion matrix allows authors to know the “true positive”, meaning that the algorithm predicts “*Yes*” while the real result is also “*Yes*”. The “true negative” in the case where the model predicts “*No*” where the result is actually “*No*”. “The false positive” when the model predicts Yes, but the actual result is No. And “the false negative” when expected No, but the result is Yes.

III. Scope Of The Study

This research is part of the first research carried out in this area in DR Congo to resolve a real problem related to the attendance of agents in the Kwilu-DR Congo public civil service. And among the numerous works carried out in the field of facial recognition using artificial intelligence and computer vision, more particularly the use of convolutional neural networks, the system proposed in this work considerably overcomes the constraint linked in the light quality of recognition, which remains a problem raised by the search of (Teoh K., 2021).

IV. Result And Discussion

Result

The implemented system, as described in the methodology, began with the development of a learning model based on a convolutional neural network, which was trained using a dataset of photographs of public sector agents. The model was evaluated, saved, and later utilized to create the automatic attendance monitoring system. The model was evaluated first by the classification report, then by the confusion matrix, and finally by the accuracy.

Figure 8 shows the evaluation of the training model based on the classification report, in which the model has a prediction capacity of 100%, an accuracy of 100%, and a specificity of 100% for all five classes of images studied.

```
In [17]: print(classification_report(trainy, yhat_train))
              precision    recall  f1-score   support

           0       1.00      1.00      1.00         4
           1       1.00      1.00      1.00         5
           2       1.00      1.00      1.00         2
           3       1.00      1.00      1.00         2
           4       1.00      1.00      1.00         5

    accuracy                           1.00        18
   macro avg       1.00      1.00      1.00        18
weighted avg       1.00      1.00      1.00        18
```

Figure 9: Model Classification Report

The confusion matrix after training the model expressing the prediction result by comparing the properties of an actual image to the predicted one as follows:

```
In [16]: print(confusion_matrix(trainy, yhat_train))

[[4 0 0 0 0]
 [0 5 0 0 0]
 [0 0 2 0 0]
 [0 0 0 2 0]
 [0 0 0 0 5]]
```

Figure 10: Confusion matrix

The confusion matrix in Figure 9 demonstrates that the predictions were accurate for each sample, depending on the number of persons. For example, the first sample of four people had four right predictions. In addition, the learning rate is 100%, as illustrated in Figure 10 below:

```
In [15]: # predict
         yhat_train = clf.predict(newTrainX)
         #yhat_test = clf.predict(newTestX)
         # score
         score_train = accuracy_score(trainy, yhat_train)
         #score_test = accuracy_score(testy, yhat_test)
         # summarize
         print('Accuracy: train=%.3f' % (score_train*100))
         #print('Accuracy: train=%.3f, test=%.3f' % (score_train*100, score_test*100))

         Accuracy: train=100.000
```

Figure 11: Model learning rate

The model was saved as shown in Figure 11 below for being used in the computer vision-based system implemented in PyCharm Software.

```
In [1]: joblib.dump(clf, "face_recognition.pkl")
```

Figure 12: Saving the model

The recorded model was then tested in the system to identify the agents using facial recognition from a camera image, and the dates and times of arrival and departure were recorded in an Excel sheet in accordance with the organization's service times. Several experiments were conducted for this purpose, and the findings were conclusive and presented. The file recording agent attendance contains three columns: the agent's name, date, and time.

Table 1: Attendance file template

Name	Date	Time

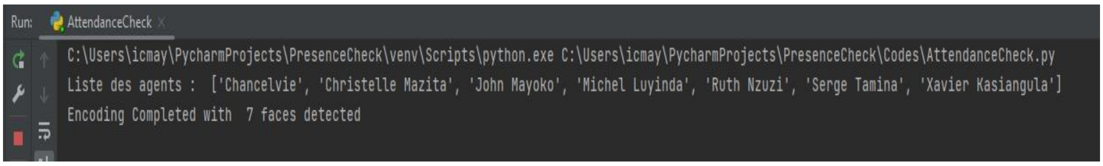
The method developed employs a model trained for the task of detecting and identifying company workers. The model was imported in the following format:

```
import cv2
import numpy as np
import face_recognition
import os
from datetime import datetime

path = 'C:\\Users\\icmay\\PycharmProjects\\PresenceCheck\\trainData\\'
```

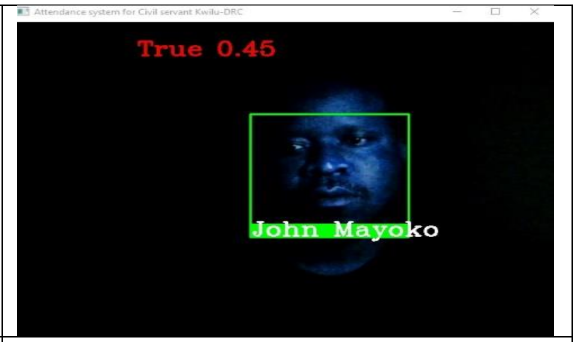
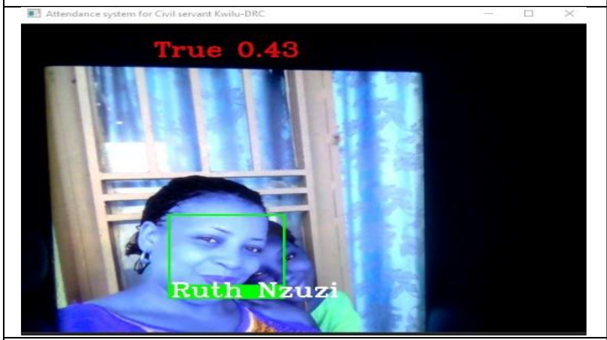
Figure 13: Importing the model and other libraries

When the intelligent attendance monitoring system is launched, an attendance file is created if it does not already exist, and the agents' names are loaded based on their corresponding photos.

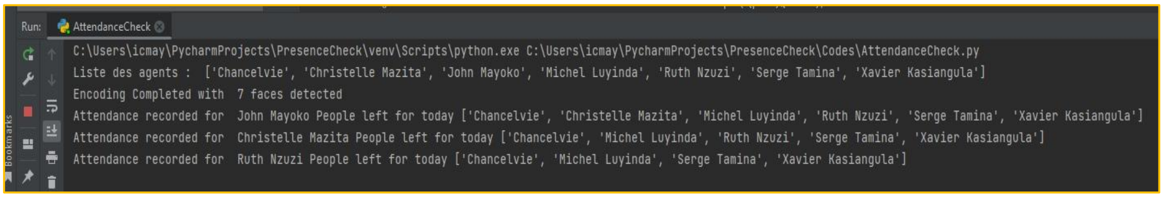


And the camera is then loaded, launched and ready for use in the agent detection and recognition process.

Table 2: Examples of agent attendance scores

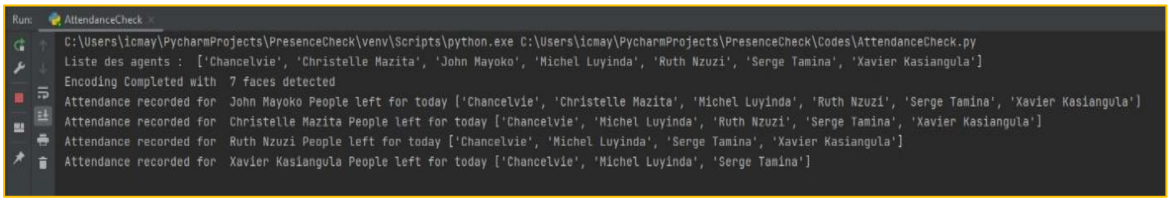
<table>
<tr><td></td><td></td></tr>
<tr><td>a) The camera was used in the dark and the system had difficulty detecting the face and performed less recognition</td><td>b) When the light was added, the system successfully detected the face and recognized the agent and the attendance was reported while specifying the agents whose attendance are not yet monitored.</td></tr>
<tr><td></td><td></td></tr>
<tr><td>Attendance for Ruth Nzuzi (c)</td><td>Attendance for Christelle Mazita (d)</td></tr>
<tr><td></td><td></td></tr>
<tr><td>Attendance for Xavier Kasiangula (e)</td><td>Attendance for Serge Tamina (f)</td></tr>
</table>

a & b

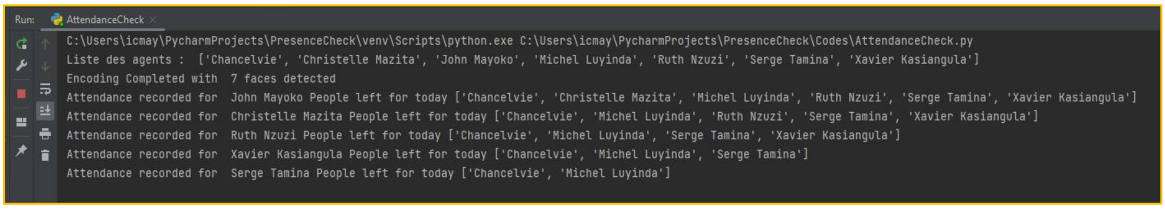


Attendance for Ruth Nzuzi (c)

Attendance for Christelle Mazita (d)



Attendance for Xavier Kasiangula (e)



Attendance for Serge Tamina (f)

When the system detects the agent, depending on the detection time, a welcome or goodbye message is presented, and the time is recorded in the little Excel database, which after a few tests appears as follows in Table 3:

Table 3: Presence of agents

Name	Date	Time
John Mayoko	14/12/2023	06-22-00
Christelle Mazita	14/12/2023	06-24-50
Ruth Nzuzi	14/12/2023	06-27-39
Xavier Kasiangula	14/12/2023	06-29-19
Serge Tamina	14/12/2023	06-31-26

Discussion

Based on the results, we can deduce that the trained model had a very good learning accuracy rate, even though the dataset did not contain enough photos per individual for learning. According to (Tesema S., 2022), a huge dataset is quite important for training object detection for performance. Despite the limited dataset volume, the model performed wonderfully. The improvement of the learning model using a large dataset of millions of images enhances detection precision so that even when the image or video provided is a little fuzzy, precise recognition remains a challenge. To do this, solutions such as utilizing the GAN approach (Generative Adversarial Network) to obtain produced images to enhance the dataset's size are encouraged.

In addition, light is an important factor that influences the detection accuracy, when the camera image is blurred, the detection accuracy is less and the recognition system tends to perform false recognition when the intensity light is weak. This could potentially be fixed by adding more training images captured with low light intensity for model training or the face classifier. Furthermore, compared to the research conducted by (Teoh K., 2021), our model gave the best result when using a camera or a simple image because the model was tested even in a dark environment to check the detection quality as shown in Table 2, image (b) in the result section.

In addition, regarding image preprocessing during model training, a Python image package called Skimage containing most of the techniques like Random Walker Algorithm (RWA), Median Filter, and the Otsu algorithm was used. An additional factor that significantly influences the recognition process throughout the learning stage.

This system can be used to detect the presence of agents for attendance marking as long as the camera is positioned at a good angle with adequate light and intensity to ensure high recognition accuracy.

V. Conclusion

This study enabled us to train a convolutional neural network model using the photos in the dataset, with the goal of facilitating the highly precise recognition of persons and automatically notifying their attendance.

Regardless of the number of photos used, the outcome of the trained model produces excellent results, allowing it to be used in the open CV library. The developed system works properly, and if it is deployed in public civil service, it will significantly improve productivity by encouraging employees to arrive on time every day rather than leaving early, in contrast to the existing system, which allows employees to arrive whenever they want or alternatively, do not come to work while attendance is checked, and this has a very negative impact on staff performance.

Furthermore, the proposed system offers an effective constraint mechanism to overcome the problem and guarantee good performance because it is not possible for one person to clock in for another or to escape work before time. This research is part of solving a real problem in society and is of great interest in the context of the reforms proposed by the Congolese government in general and that of the Kwilu province in particular.

Furthermore, the suggested system provides an effective constraint mechanism to address the issue and ensure good performance because nobody can mark attendance another or leave work early. Therewith, this research fits into the framework of solving a real societal problem and is highly relevant to the reforms that the Congolese government, in general, and the Kwilu Province in particular, have proposed.

Declarations

Acknowledgments: The authors would like to thank the public civil service of the Bandundu-DR Congo for the cooperation in the data research, as well as the Computer Science Section of Grande Ecole des Commerces, Bandundu for the support given. In addition, would also thank the University of Kikwit, the University of Science and Technology of Kinshasa (USCITECH), the Machine Learning and Optimization Research Laboratory (Optimall) for making this research possible.

Author Contributions: Jean Chrysostome Mayoko Biong - Definition of intellectual content, Literature survey, Prepared first draft of manuscript, implementation of study protocol, data collection, data analysis, manuscript preparation, and manuscript revision; Aser Itakala and Espérant Kikata – model implementation; Chadrack Lubamba Milolo, Serge Tamina Moke, Aser Itakala and Hervé Lamanabwe - Concept, design, data collection, model testing and result evaluation, manuscript preparation, and manuscript revision; Slife Nyazabe - Concept, design, literature survey; Jean Chrysostome Mayoko Biong and Slife Nyazabe - Review manuscript, Jean Chrysostome Mayoko Biong - submission of article.

Conflict of Interest: The authors declare no conflict of interest.

Consent to Publish: The authors agree to publish the paper in International Journal of Recent Innovations in Academic Research.

Data Availability Statement: Data are contained within the article.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Research Content: The research content of manuscript is original and has not been published elsewhere.

Reference

[1] Abdulqader M. (2021). Deepfakes Detection Techniques Using Deep Learning: A Survey. Journal Of Computers And Communications , 20-35. Retrieved From Https://Www.Scirp.Org/Journal/Paperinformation.Aspx?Paperid=109149

[2] Bishop M. (1995). Neural Networks For Pattern Recognition. Oxford University Press , 1-181.

[3] Chartrand G., Pm (2017). Deep Learning: A Primer For Radiologists. Radiographics , 113-2131.
[4] Dimitrovsky I., Ki (2023). Current Trends In Deep Learning For Earth Observation: An Open-Source Benchmark Arena For Image Classification. Isprs Journal Of Photogrammetry And Remote Sensing , 18-35.
[5] Erickson Bj., Kp (2017). Machine Learning For Medical Imaging. Radiographic , 505-515. Retrieved From Https://Pubs.Rsna.Org/Doi/10.1148/Rg.2017160130
[6] Feng A., Xy (2023). Label Propagation Via Random Walk For Training Robust Thalamus Nuclei Parcellation Model From Noisy Annotations. Arxiv , 1-5.
[7] Feng A., Xy (2023). Label Propagation Via Random Walk For Training Robust Thalamus Nuclei Parcellation Model From Noisy Annotations. Arxiv , 1-5.
[8] Grady L. (2006). Random Walks For Image Segmentation. Ieee Transcations On Pattern Analysis And Machine Intelligence , 1-10.
[9] Hima Bindu Ch. And Satya Prasad K. (2012). An Efficient Medical Image Segmentation Using Conventional Otsu Method. International Journal Of Advanced Science And Technology , 1-8.
[10] Jassim Fa And Altaani Fh, J. (2013). Hybridization Of Otsu Method And Median Filter For Color Image Segmentation. International Journal Of Soft Computing And Engineering (Ijsce) , 1-6.
[11] Kharwal A. (2021, 07 07). Classification Report In Machine Learning . Retrieved From Https://Thecleverprogrammer.Com/: Https://Thecleverprogrammer.Com/2021/07/07/Classification-Report-In-Machine-Learning/
[12] Khouja, Ru (2015). Compression Of Convolutional Neural Networks: The Contribution Of Regularized Cp Decomposition. Https://Gretsi.Fr/Data/Colloque/Pdf/2023_Khouja1359.Pdf , 1-4.
[13] Kim B. And Lee J. (2019). A Video-Based Fire Detection Using Deep Learning Models. Applied Sciences , 1-19. Retrieved From Https://Www.Mdpi.Com/2076-3417/9/14/2862
[14] Krizhevsky, A.S. (2017). Imagenet Classification With Deep Convolutional Neural Networks. Communications Of The Acm , 84-90.
[15] Kumar A., Cs (2022). Face Detection And Recognition Using Opencv. International Journal Of Computer Applications , 23-32.
[16] Long J., Se (2015). Fully Convolutional Networks For Semantic Segmentation. Ieee Conference On Computer Vision And Pattern Recognition (Cvpr), Los Alamitos, Ca, Usa, Jun 2015 , 3431–3440.
[17] Makkar H. And Pundir A. (2014). Image Analysis Using Improved Otsu's Thresholding Method. International Journal On Recent And Innovation Trends In Computing And Communication , 2122 – 2126.
[18] Masakuna Jf (2019). Machine Learning Methods For Shark Detection. Arxiv , 1-31.
[19] Matuska S., Hr (2012). The Comparison Of Cpu Time Consumption For Image Processing Algorithm In Matlab And Opencv. Ieee Xplore , 75–78.
[20] Mohammed Zeeshan M. (2020, 04 26). Cost, Activation, Loss Function|| Neural Network|| Deep Learning. What Are These? Retrieved From Https://Medium.Com/: Https://Medium.Com/@Zeeshanmulla/Cost-Activation-Loss-Function-Neural-Network-Deep-Learning-What-Are-These-91167825a4de
[21] Nielsen M. (2023, 11 29). Neural Networks And Deep Learning . Retrieved From Http://Neuralnetworksanddeeplearning.Com: Http://Neuralnetworksanddeeplearning.Com/Index.Html
[22] Nikitha P. Et Al. (2023). Object Detection Using Opencv With Python. International Research Journal Of Modernization In Engineering Technology And Science , 1-6.
[23] Rakotonirina H., Hp (2023). Spatial Interpolation With A Generative Neural Network As An Alternative To Kriging. Https://Gretsi.Fr/Data/Colloque/Pdf/2023_Rakotonirina1273.Pdf , 1-4.
[24] Ren J. And Wang Y. (2022). Overview Of Object Detection Algorithms Using Convolutional Neural Networks. Journal Of Computers And Communications , 115-132. Retrieved From Https://Www.Scirp.Org/Journal/Paperinformation.Aspx?Paperid=115011
[25] Rumelhart Rj And Hinton Ge (1986). Learning Internal Representations By Error Propagation In Parallel Distributed Processing: Explorations In The Microstructure Of Cognition. Foundations. Mit Press .
[26] Saidi R. (2023). Facial Recognition By Deep Learning. Dspace , 1-81. Retrieved From Https://Dspace.Univ-Ouargla.Dz/Jspui/Bitstream/123456789/34752/1/Saidi.Pdf
[27] Simonyan K., A. Z. (2014). Very Deep Convolutional Networks For Large-Scale Image Recognition. Arxiv Preprint:1409.1556 , 1-14.
[28] Sriratana, W.M. (2018). Application Of The Opencv-Python For Personal Identifier Statement. International Conference On Engineering, Applied Sciences, And Technology (Iceast) , 1-4.
[29] Teoh K., Ir (2021). Face Recognition And Identification Using Deep Learning Approach. Journal Of Physics: Conference Series , 1-10. Retrieved From Https://Iopscience.Iop.Org/Article/10.1088/1742-6596/1755/1/012006/Pdf
[30] Tesema S. (2022). Accelerating Deep Convolutional Neural Network-Based Object Detection Inference Using Fpga. Theses.Fr , 1-125.
[31] Yang X., S.X. (2012). An Improved Median-Based Otsu Image Thresholding Algorithm. Aasri Procedia , 468-473.
[32] Zhan Y. And Zhang G. (2019). An Improved Otsu Algorithm Using Histogram Accumulation Moment For Ore Segmentation. Symmetry , 1-15. Retrieved From Https://Www.Mdpi.Com/2073-8994/11/3/431