A Moroccan Sign Language Recognition Algorithm Using a Convolution Neural Network

Nourdine Herbaz*, Hassan El Idrissi and Abdelmajid Badri

Laboratory of Electronics, Energy, Automation & Information Processing (LEEA&TI), Faculty of Sciences and Techniques Mohammedia, Hassan II University of Casablanca, 28830, Morocco
E-mail: herbaznourdine@gmail.com; idrissi99@yahoo.fr; abdelmajid_badri@yahoo.fr
**Corresponding Author*

Received 16 March 2022; Accepted 11 April 2022;
Publication 10 August 2022;
Withdrawn 01 September 2022; Resubmitted 20 September 2022;
Republished 20 October 2022

Abstract

Gesture recognition is an open phenomenon in computer vision, and one of the topics of current interest. Gesture recognition has many applications in the interpretation of sign language, one is in human–computer interaction, and the other is in immersive game technology.

For this reason, we have developed a model of image processing recognition of gestures, based on artificial neural networks, starting from data collection, identification, tracking and classification of gestures, to the display of the obtained results. We propose an approach to contribute to the translation of sign language into voice/text format.

In this paper, we present a Moroccan sign language recognition system using a convolutional neural network (CNN). This system includes an important data set of more than 20 files. Each file contains 1000 static images of each signal from several different angles that we collected with

Journal of ICT Standardization, Vol. 10_3, 411–426.
doi: 10.13052/jicts2245-800X.1033
© 2022 River Publishers

a camera. Different sign language models were evaluated and compared with the proposed CNN model. The proposed system achieved an accuracy of 99.33% and achieved best performance with an accuracy rate of 98.7%.

Keywords: Sign language, convolutional neural networks, deaf people, image processing, real time.

1 Introduction

For many years, gestures have been a basic method of communication. The main objective is to recognize these gestures with the help of innovative technologies. At present, the number of deaf people who use a signed language, such as the Moroccan Sign Language, as their preferred means of communication, is very large [1]. It would be very interesting for people who do not use this language to understand it. Despite the achievements of modern science in this direction, sign language remains one of the least studied areas [1]. Modern advances in the field of neural networks and deep learning, as well as the increase in computing power, allow for the development of innovative solutions that help people who are hard of hearing with their everyday lives. To solve this problem, in our research, we propose a convolutional neural network (CNN) and a very important dataset for image classification. Figure 1 shows some hand gestures of Arabic Sign Language Alphabets with the meaning of each gesture. Our goal is to translate the sign language into text.

In this paper, we present an approach to improve the performance of hand gesture recognition using a camera. The document is composed of five sections. In Section 2, we start with an overview of related works and the methodology we propose for our solution is presented in Section 3. Section

Figure 1 Arabic sign language alphabet [2].

4 focuses on the obtained results. In Section 5 we end our paper with a conclusion and perspectives.

2 Related Work

In the light of rapid changes in technology and the varying conditions in understanding and studying languages, sign language faces many challenges. Many researchers have taken on this challenge by developing methods to understand sign language symbols. In principle, we say that sign language is an obstacle when non-signers try to use it to speak.

Work on hand gesture recognition was originally seen as a method of human–computer interaction, with [3] exploring three different gesture feature extraction applications, scale invariant feature transform (SIFT), local binary pattern (LBP), and histogram of oriented gradients (HOG). Most applications have been developed on static gesture recognition, i.e., geometric feature extraction, the number of fingers lifted in a gesture, the distance from the center of the palm of the hand to the fingertips and the valley between the fingers [4]. The method presented in [5] lifted some of these restrictions, but required data augmentation to improve the accuracy of deep CNNs. Binary image methods are often based on fitting "HU moments" for feature extraction and classification using the KNN algorithm "HU moments" (e.g., [8]) and show excellent accuracy. The effect of hyper-parameters on each hand gesture presented in [6] is also one of the means used by the authors to decode hand gestures from EMG data recorded in 18 subjects regarding hand gestures using the implementation of convolutional neural networks. Sidig et al. [7] presented a model to collect a database for ArSL using Microsoft Kinect-V2.

With the technological revolution and artificial intelligence, the recognition of sign language based on CNNs has become the main focus of research, giving rise to many applications based on CNNs. The feature extraction technique oriented FAST and rotated BRIEF (ORB) [9] used many different pre-treatment techniques such as a gradient histogram, LBP and principal component analysis (PCA) on the same dataset, then it uses a support vector machine to classify them. The method in [11] uses a set of geometric characteristics of the hand silhouette based on a Fisher Vector compact representation to define the hand posture. In [15], the authors used a support vector machine and a CNN to detect and recognize American sign language; the authors achieved an accuracy of 98.5%.

3 Methodology

The proposed static gesture recognition methodology consists of three phases, as shown in Figure 2. These three phases deal with hand gesture collection, image training, classification and gesture detection. The hand area is recognized from image processing using the transformation of an RGB image to a binary image. The resulting hand image is then segmented into two areas: a white area consisting of the palm and fingers and a black area consisting of the outer area of the hand, which is the hand's environment. The black area does not play any role in distinguishing the different gestures. The following features are extracted: orientation of the hand and number of fingers present in the gesture. Finally, the extracted features are given to the input of a CNN sequence for gesture classification, as shown in Figure 3.

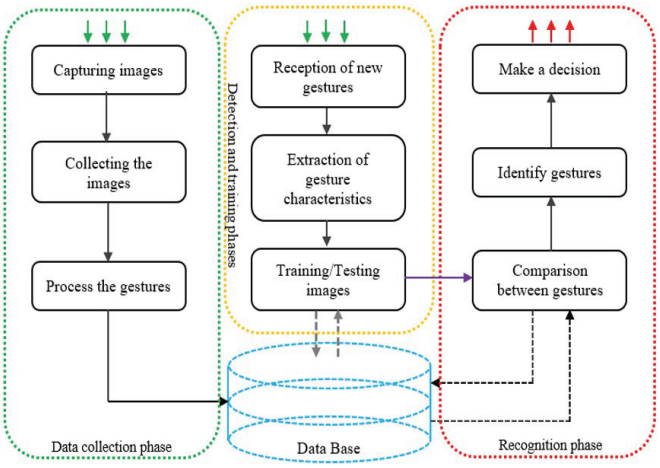


Figure 2 Block diagram of the system.

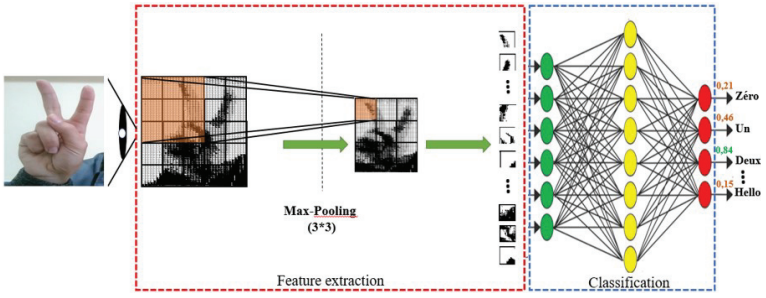


Figure 3 A CNN sequence to classify gestures.

3.1 System Architecture

Our classification architecture is classical, combining convolution and max pooling. However, to obtain a fast classification allowing real-time classification and localization, we have chosen a lightweight network. Figure 4 shows the six layers of our convolutional network: the first layers are used to identify simple shapes (vertical, horizontal, oblique lines, etc.); the next ones pool this information to recognize more complex shapes; and the last one, called a "classifier", is able to classify information and make a "decision" (our example, "gesture recognition"). Finally, the extracted features are given to the input of a CNN sequence for gesture classification.

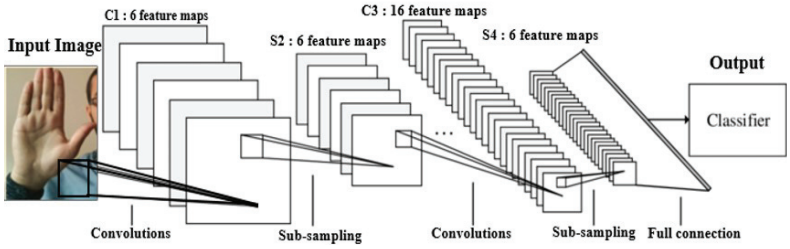


Figure 4 Architecture of a CNN.

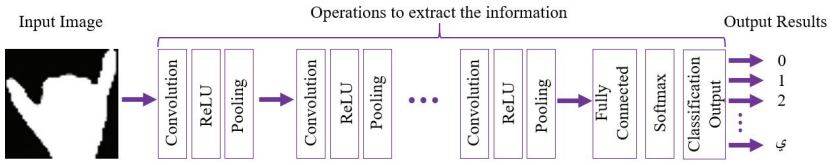


Figure 5 CNN processus.

To recognize the hand gesture we used a CNN, Figure 5 shows the CNN processus. We have as input an original image of hand gesture, which will go through many operations (convolution, ReLU, pooling, etc.) to extract the information of the sign language to make a decision in output.

3.2 Dataset

To evaluate our system, we used a database containing hand gestures with meanings. Figure 6 represents some images from our database before converting them into binary.

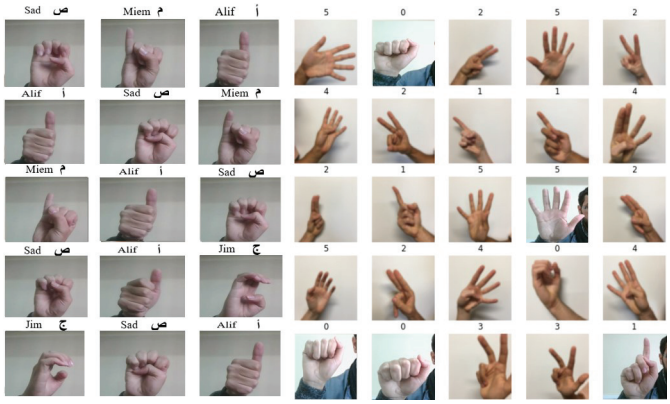


Figure 6 Images of the gestures in our database.

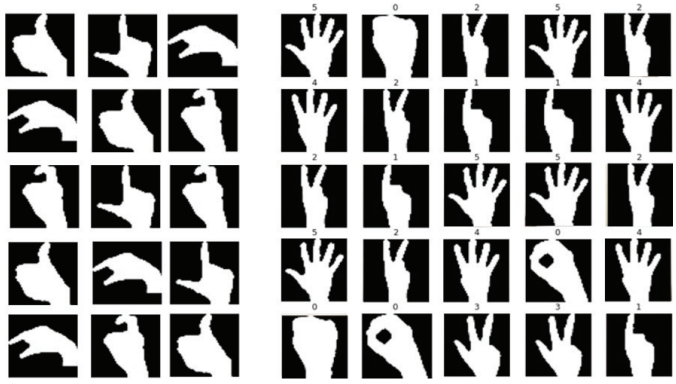


Figure 7 Gesture images from our database converted to binary.

3.3 Equations

In each evolution of the image during a CNN sequence, our objective is to reduce the size of the selected image, as shown in Figure 3, and for this we need to reduce the dimensions of the output feature map, relative to the input, using the following general formula (1):

$$\text{Size of the future image} = \left(\frac{n+2p-f}{s}+1\right)\left(\frac{n+2p-f}{s}+1\right). \quad (1)$$

with,

n: size of the input image
p: padding amount

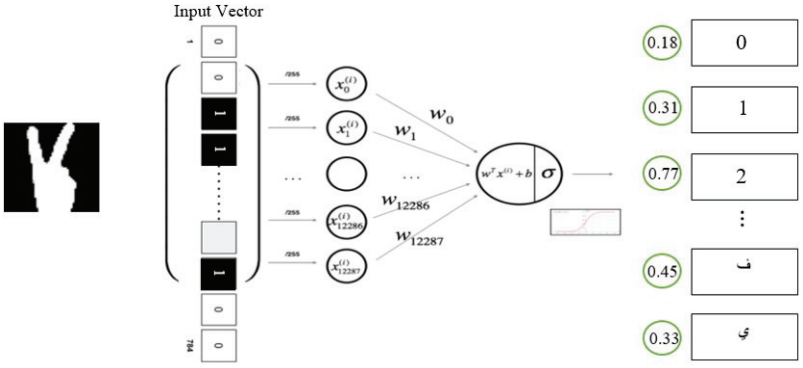


Figure 8 Vector transformed binary image for neural network system adaptation.

f: filter size
s : stride (number of pixels shifted on the input image).

In Figure 8 we use in the input layer the values calculated previously, we then add from 1 to an infinity of hidden layers to our network. Finally, we add in the output layer the number of adequate neurons (e.g. 20 neurons if we wish our network to predict the probability that the image belongs to gesture 1, gesture 2. . . or gesture 20).

$$Z_i = x_i * W_i + b. \tag{2}$$

Z_i is the input of the layer i. It is performed by remembering long-term information and deciding which value to pass on to the other time step block, with W_i and b being the weights and bias parameters associated with each Z_i gate. These parameters are used to minimize the error of our training data, as shown in Equations (2) and (3).

$$y(i) = a(i) = \sigma(w^T x + b) = \frac{1}{1 + \mathrm{e}^{-(\mathrm{w^T x + b})}}. \tag{3}$$

The sigmoid function $\sigma(Z)$ is used as an activation function in neural networks. A weighted sum of inputs passes through an activation function and this output serves as input to the next layer, as shown in Equation (4).

$$\sigma(Z) = \frac{1}{1 + \mathrm{e}^{-(\mathrm{Z})}}. \tag{4}$$

In our solution, there are generally four evaluation measures used [13,14]. These are accuracy, recall, precision and F-measure. The confusion matrices corresponding to false negative (FN), true negative (TN), true positive (TP)

and false positive (FP) are used to measure the above parameters of the classifier. The following formulas define each of the metrics:

TP: is a positive example that correctly predicts the positive class.
TN: refers to the negative that is properly characterized as negative.
FP: refers to a negative example that is incorrectly characterized as positive.
FN: refers to the positive example that is imperfectly characterized as negative.
Precision: precision is a measure of the accuracy of detection.
Accuracy: accuracy defines the ratio of predicted positives that are real positive.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{5}$$

Recall: recall defines that the number of positives recognized correctly.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{6}$$

F-measurement: the F-measurement defines the voice means of recall and precision.It is calculated by the formula below which takes both false positives and false negatives into account using the following Equation (7):

$$\text{F-measure} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}. \tag{7}$$

Accuracy: accuracy is a measure of the accuracy of the detection.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \tag{8}$$

3.4 Flow Chart

The first step is to capture the images from the webcam. The video is divided into images (300 images per second), each of which is converted into a binary image. In addition, the transformation of the images into binary is aligned to normalize some parts of the noisy images, as shown in Figure 9. After the transformations, the hand in the image is searched for using gesture detection, and then the gesture features are detected to further establish the type and orientation of the gestures. After training, the system can recognize hand gestures. Figure 9 shows the working process of the hand gesture recognition system for interpreting sign language.

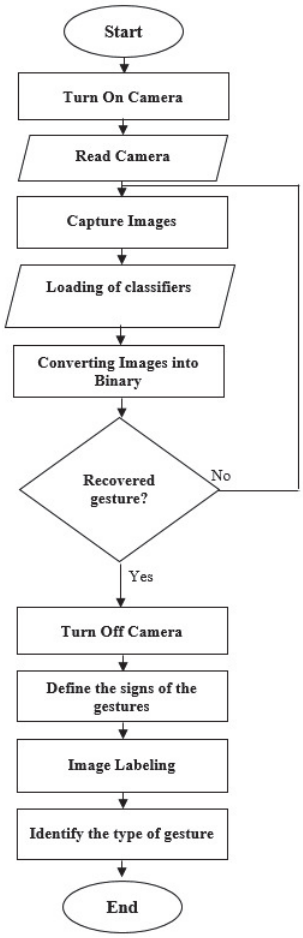


Figure 9 Flow chart of the gesture recognition model.

4 Results

A total of 20 signs have been stored in our database, which has been carefully tested using our convolution model. The system successfully takes signs as input and translates them each to their meaning as a letter or number.

The input image has been converted from an RGB color space to a grayscale image. The first image with noise and then configured to obtain a clear image, through which the hand gesture is extracted from the binary image and used for further processing, as shown in Figure 10.

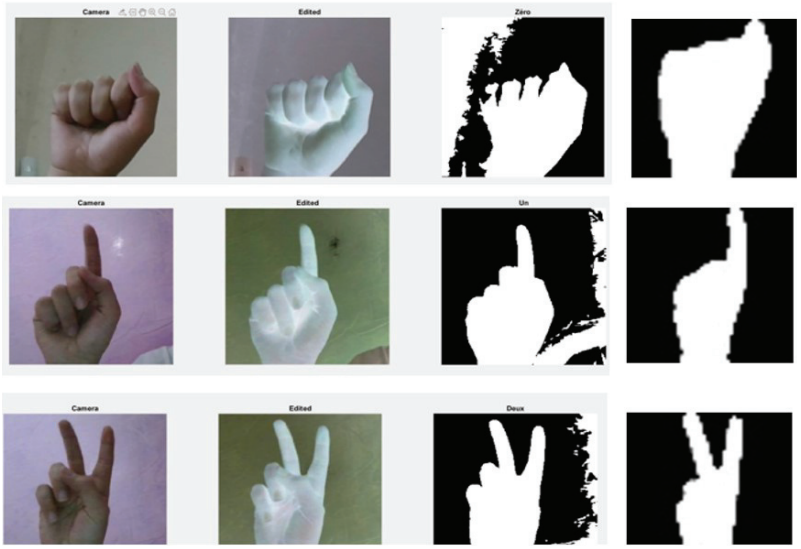


Figure 10 Color image converted to binary via a grayscale image, binary image with noise and filtered binary image.

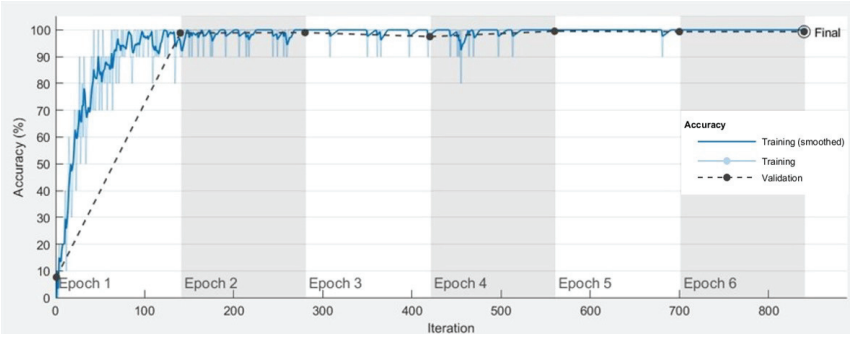


Figure 11 Accuracy performance.

After six iterations, the accuracy value increased. Training accuracy is the value of calculating the accuracy of the training dataset and predictions from the model, as shown in Figure 11.

In Figure 12, we see after six iterations that the loss value decreases because it is minimized by the convolution neural network. Based on the above loss model chart, it can be seen that training loss keeps decreasing with increasing interations.

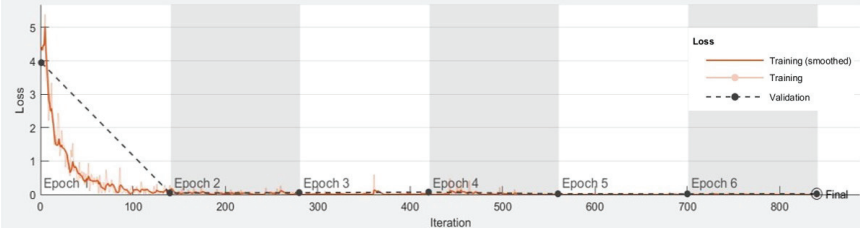


Figure 12 The loss function.

Table 1 Comparison between our proposed model and other methods

Methods	No. of Gesture	Accuracy (%)
Gaussien SVM [4]	10	95.7
Double layer CNN [15]	25	98.5
Our CNN model	**20**	**98.7**

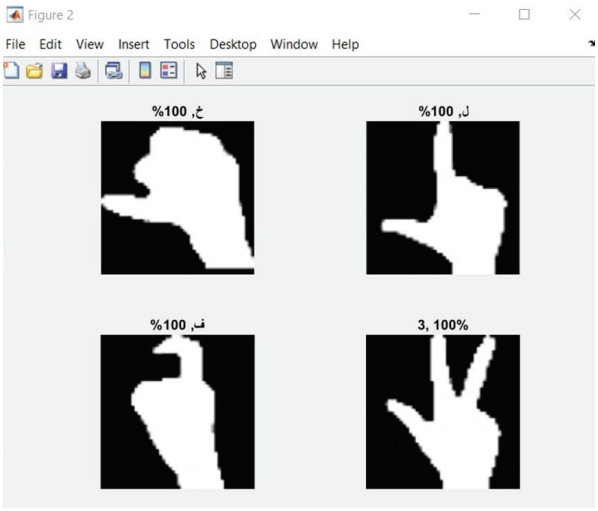


Figure 13 Final results with the accuracy rate of each gesture.

Table 1 gives the comparison results for our model with a other state-of-the-art models. It should be noted that if we use a poor dataset, the classification accuracy will be reduced, thus achieving an unsatisfactory result.

The system successfully takes signs as input and translates them each to their meaning as a letter or number with a rate of learning that is very important, as shown in Figure 13.

5 Conclusion

This paper has realized a solution for real-time detection of hand gestures captured by our camera. In order to improve the efficiency and accuracy, we use convolutional neural networks to process and extract the meaning of the gesture. Finally, we compared our results with related works. The results obtained by our CNN model are satisfactory. Our method has achieved the best performance with an accuracy rate of 98.7%.

In our view, we can improve our solution to build understandable sentences and send them as text or speech.

6 Future Directions

Sign language gestures to communicate are incomprehensible to non-signing people. Our main objective is to contribute to helping the "non-signing people" to decode and interpret sign language, using artificial intelligence techniques. We are thinking of developing an application performing the desired interpretation, from static or dynamic images taken using a camera, or from information received from flex sensors implanted on a glove. Optimization and interpretation in real time remains a considerable asset.

References

[1] S.M. Pascua; P.L.C. Espina; R.P.L. Talag; L.N. Villegas; L. Aquino De Guzman. A Filipino Sign Language Thesaurus Management System Using Ren-py. *IFLA WLIC 2017 – Wrocław, Poland – Libraries. Solidarity. Society.*, 2017.

[2] N. El-Bendary; H. Zawbaa; M. Daoud;A.E. Hassanien; K. Nakamatsu. *International Journal of Computer Information Systems and Industrial Management Applications*, 590–595, 2010.

[3] F. Zhang. Human-Computer Interactive Gesture Feature Capture and Recognition in Virtual Reality. *Ergonomics in Design*, 29(2):19–25, 2021.

[4] P. Sharma and A.R. Shyam. Depth data and fusion of feature descriptors for static gesture recognition. *IET Image Processing*, 14(5): 909–920, 2020.

[5] Q. Zheng; M. Yang; X. Tian; N. Jiang and D. Wang. A Full Stage Data Augmentation Method in Deep Convolutional Neural Network for Natural Image Classification. *Discret. Dyn. Nat. Soc*, 1(11), 2020.

[6] A.R. Asif; A. Waris; S.O. Gilani; M. Jamil; H. Ashraf; M. Shafique; I.K. Niazi. Performance Evaluation of Convolutional Neural Network for Hand Gesture Recognition Using EMG. *Sensors*, 20(6), 2020.

[7] A. A. I. Sidig; H. Luqman; S. Mahmoud; M. Mohandes. KArSL: Arabic Sign Language Database *ACM Transactions on Asian and Low-Resource Language Information Processing*, 20(1), 1—19, 2021.

[8] R. Nair; K.A. Dileep; Ashu; S. Yadav; B. Sourabh. Hand Gesture Recognition system for physically challenged people using IoT. *6th International Conference on Advanced Computing & Communication Systems (ICACCS)*, 671–675, 2020.

[9] A. Sharma; A. Mittal; S. Singh; V. Awatramani. Hand Gesture Recognition using Image Processing and Feature Extraction Techniques. *Procedia Comput. Sci*, 173:181—190, 2020.

[10] J.P. Sahoo; A.J. Prakash; P. Pławiak; S. Samantray. Real-Time Hand Gesture Recognition Using Fine-Tuned Convolutional Neural Network. *Sensors*, 22(3), 706, 2022.

[11] L. Fang; N. Liang; W. Kang; Z. Wang; D.D. Feng. Real-time hand posture recognition using hand geometric features and Fisher Vector. *Signal Processing: Image Communication*, 82, p:115729, 2019.

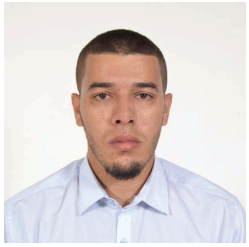
[12] Y.S. Tan; K.M. Lim; C.P. Lee. Hand gesture recognition via enhanced densely connected convolutional neural network. *Expert Syst. Appl.*, 175, p:114797, 2021.

[13] C. Arun; R. Gopikakumari. UOptimisation of both classifier and fusion-based feature set for static American sign language recognition. *IET Image Process.*, 14(10):2101–2109, 2020.

[14] X. Tang; Z. Yan; J. Peng; B. Hao; H. Wang; J. Li. Selective spatiotemporal features learning for dynamic gesture recognition. *Expert Syst. Appl.*, 169, p:114499, 2021.

[15] V. Jain; A. Jain; A. Chauhan; S.S. Kotla; A. Gautam. American Sign Language recognition using Support Vector Machine and Convolutional Neural Network. *International Journal of Information Technology.*.13(3):1193—1200, 2021.

Biographies



Nourdine Herbaz is a Ph.D. student in Electronics, Energy, Automation & Information Processing (LEEA&TI) laboratory, Faculty of Sciences and Techniques of Mohammedia, Hassan II University of Casablanca, Morocco. He did his bachelor's degree in Biomedical Instrumentation and Maintenance from Hassan I University in 2016 from Morocco, and Masters' degree in Embedded systems and Mobile from Tunisia. His research interest is Artificial Intelligence, Neural Network, Embedded systems... Currently, the project on which he is working is focused on applications of artificial intelligence for the interpretation of sign language.

Hassan El Idrissi is a full professor since 1994 in the electrical engineering department at the Faculty of Sciences and Techniques of Mohammedia in Hassan II University Casablanca – Morocco, where he teaches courses on the physics of semiconductors, sensors, electronics and graphic programming dedicated to instrumentation, automation, and supervision. His thesis defended in 1993 at the Institute of Electronics and Microelectronics of the North, of the University of Science and Technology of Lille in France, focused on field effect transistors with insulated gate. He has supervised theses in the field of semiconductors and magnetic pulse generators. He has

participated in several national and international congresses and conferences. His current research focuses on artificial intelligence and its societal applications, more particularly the electronic coding of sign language by camera or smart glove.

Abdelmajid Badri is a holder of a doctorate in Electronics and Image Processing in 1992 at the University of Poitiers–France. In 1996, he obtained the diploma of the authorization to Manage Researches (Habilitation à Diriger des Recherches: HDR) to the University of Poitiers–France, on the image processing. Qualified by the CNU-France in 61th section (informatics Engineering, Automatic and Signal processing. He is an University Professor (PES-C) at the University Hassan II of Casablanca – Morocco (FSTM) where he teaches the electronics, the signal processing, image processing and telecommunication (Department of Electric Engineering). He is a member of the laboratory EEA&TI (Electronics, Electrotechnics, Automatic and information Processing) which he managed since 1996. The research works of A. Badri concerns the communication and Information Technology (Electronics Systems, Signal/Image Processing and Telecommunication). He managed several doctoral theses. He is a co-author of several national and international publications. He is responsible for several research projects financed by the ministry or by the CNRST or by the industrialists. He was member of several committees of programs of international conferences, reviewer of several revues and chairman of several international congresses in the same domain. He is a member and coresponsible in several scientific associations in touch with his domain of research. He is an expert CNRST and Ministry. He was responsible for several academic structures (Director of ESTC, Director of ENSAMC an interim, Vice Dean FSTM, Head of the Electric Engineering Department).