HANDWRITTEN DIGIT RECOGNITION SYSTEM BASED ON CNN AND SVM

Yousra BERRICH and Zouhair GUENNOUN

Smart Communications Research Team - ERSC, Mohammadia School of Engineering, Mohammed 5 University in Rabat, Morocco

ABSTRACT

The recognition of handwritten digits has aroused the interest of the scientific community and is the subject of a large number of research works thanks to its various applications. The objective of this paper is to develop a system capable of recognizing handwritten digits using a convolutional neural network (CNN) combined with machine learning approaches to ensure diversity in automatic classification tools. In this work, we propose a classification method based on deep learning, in particular the convolutional neural network for feature extraction, it is a powerful tool that has had great success in image classification, followed by the support vector machine (SVM) for higher performance. We used the dataset (MNIST), and the results obtained showed that the combination of CNN with SVM improves the performance of the model as well as the classification accuracy with a rate of 99.12%.

KEYWORDS

Classification, feature extraction, convolutional neural network, support vector machine, MNIST.

1. INTRODUCTION

The recognition of handwritten digits (RCM) is part of the Optical character recognition (OCR) among its applications: the postal sorting where every day thousands of items are automatically sorted, moreover there is the reading of the numerical amount of bank checks.

Even after several works in this direction, the recognition performance remains far from that of the human eye due to the similarity between handwritten digits or the difference between writing styles, so it is necessary to develop a system able to acquire an image of an input digit and recognize the digit present in the image. The objective of this work is to create an automatic system capable of recognizing the number of manuscripts in the MNIST database with resolution (28*28) pixels [01].

In this proposed model, we decide to verify its efficiency in recognizing the handwritten digit while using the SVM classifier at the last layer of the CNN neural network; this proposed method shows an excellent performance with a high accuracy.

The structure of this paper is articulated around three parts: the first part is devoted to give a general overview on some previous works in the direction of handwritten digit recognition, then in the second part we will present the process of the handwritten digit recognition system and at the end we will elaborate an implementation for handwritten digit recognition using a classification approach based on convolutional neural network (CNN) combined with SVM and we will present our results obtained in the form of graphs and in a summary table.

2. Some Previous Works

The recognition of the writing is known under the name of O.C.R (Optical Character Recognition). The first works date back to the 1900s by TYURIN, during which the scanning scanner for television was invented and in the year 1912 by ALBE and in the year 1925 by Thomas, which mimicked the human interpretation of visual computing, a point of transformation appeared with the invention of the first computer in 1946 by MAUCHLY and ECKERT, a few years later the first experiments in character recognition could be carried out during the sixties, and seventies, the first systems of automatic writing of the printed text were born. [02]

In addition, in 1975 the Japanese were using readers that decipher the postal code written by hand or typed. In the same period, the French company CONTER built a system of automatic reading of printed text intended for the blind, then the American company KURZWELL improved the previous system by proposing reading machines for the blind formulating the text aloud by vocal synthesis. [03]

During this phase the researchers encountered many difficulties in addition to the complexity of the problem of recognition due to the great variability of handwriting, the non-availability of memory and computing power for the realization of concrete systems operating in real time, on the other hand since 1980 when the recent electronic progress and more particularly the advent of powerful computers at low cost made it possible to solve this type of problem and the research in handwriting recognition multiplied in a spectacular way, and many new techniques were born.

3. Characterization of the Handwritten Digit Recognition System

Handwritten digit recognition systems are generally based on the following main steps:
Signal acquisition, pre-processing, feature extraction, classification.

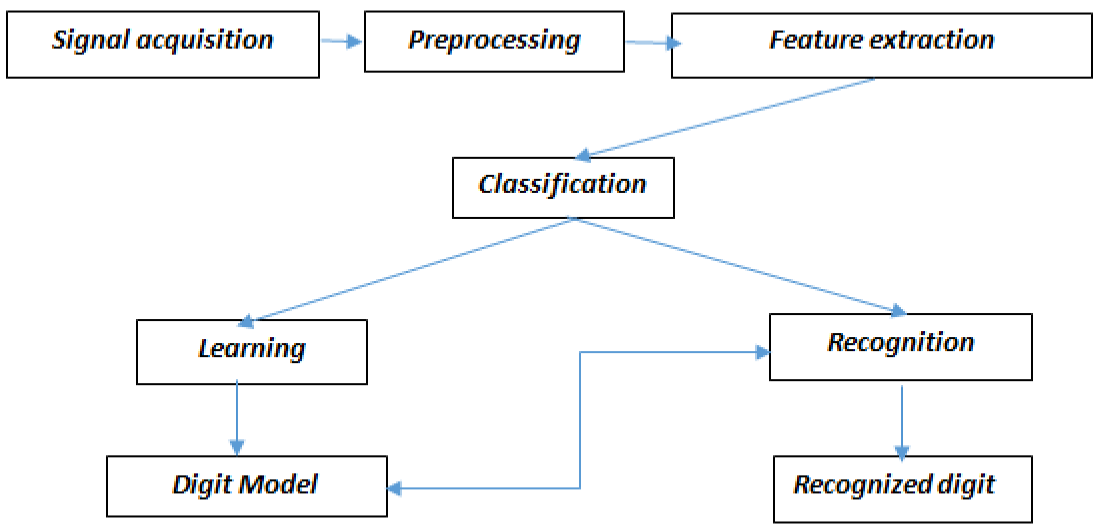

Figure 1.The process of the handwritten digit recognition system

3.1. Signal Acquisition

This step consists in converting a paper document into a digital image format with a minimum of degradation. However, despite all this, some noise may appear and cause a loss of quality due to the work area and its lighting [04].

3.2. Preprocessing

The purpose of pre-processing is to reduce the noise superimposed on the data and to normalize the image size while keeping the significant information of the presented shape. The pre-processing operations generally used are binarization, smoothing (noise), normalization (size normalization, straightening normalization), Tinning [05].

3.3. Feature Extraction

In a handwritten digit recognition system, the feature extraction phase consists of obtaining the most relevant volume of information that will be provided to the classifier later on.

3.4. Classification

Classification allows to transform the attributes characterizing the shapes into class membership (passage from the coding space to the decision space).

4. IMPLEMENTATION AND RESULTS

The goal of our work is to develop a cipher recognition system while relying on CNNs and SVM as a classifier. We developed a pre-trained model by adding the SVM classifier to the last layer of the CNN neural network, and then we added a set of images to the real input set.

4.1. Dataset

The **MNIST** database (**Modified National Institute of Standards and Technology** database) is a large collection of handwritten digits. It has a training set of 60,000 examples, and a test set of 10,000 examples. It is a subset of a larger NIST Special Database 3 (digits written by employees of the United States Census Bureau) and Special Database 1 (digits written by high school students) which contain monochrome images of handwritten digits [06].

The original black-and-white (bi-level) NIST images were normalized in size with digits centered in the middle to fit in a 20x20-pixel box while preserving their aspect ratio. The resulting images were centered in a 28x28 image keeping the aspect ratio of the image.

4.2. Augmentation Data

We adopted an image augmentation technique that consists of increasing the size of a training dataset by creating new modified versions of the images from the available training images by ensuring a variation of images that makes the model able to generalize what it has learned to new images and that strongly improves the performance of this model.

4.3. Libraries

We start by importing our libraries and dataset. The first library we import is Tensorflow [07]. It is an open source library, developed by Google and released in 2015, which is very popular in Machine Learning.

In Tensorflow, we import Keras, which is a programming interface and can decrease the development time of a neural network prototype by 30%). [08]

4.4. SVM (Support Vector Machine)

SVMs are a family of machine learning algorithms that can be used to solve classification, regression and anomaly detection problems. The SVM classifier is an algorithm that maximizes the margin between classes of the problem to be solved by a hyperplane and reduces the classification error.

The SVM classifier is a machine-learning algorithm that maximizes the margin between classes of the problem to be solved by a hyperplane and reduces the classification error. The calculation of this hyperplane is based on maximizing the margin between the closest learning examples that belong to different classes [09, 10, 11].

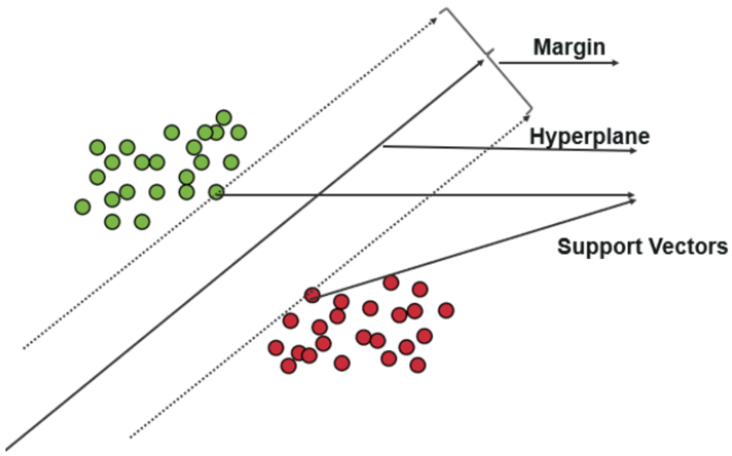

Figure 2. The classifier Support Vector Machine

4.5. Proposed Architecture

Figure 3 shows the proposed architecture principle.

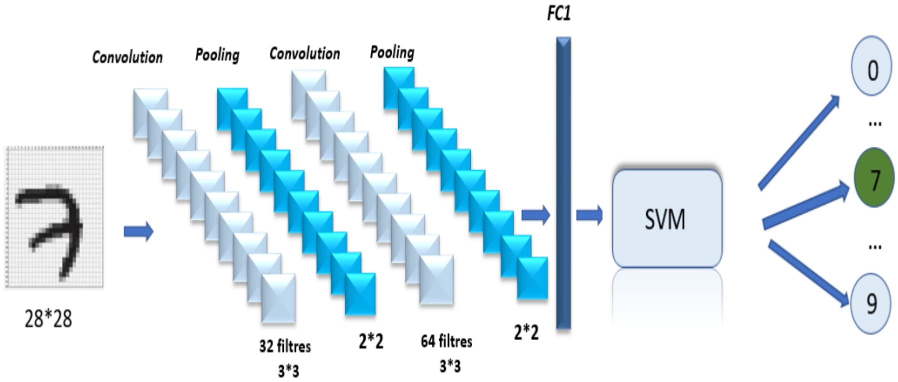

Figure 3. The proposed Architecture

The model that we present in figure 3 is composed of two convolution layers, two Maxpooling layers [12], and a fully connected layer, finally an output layer. The input image is of size 28*28, it passes to the first convolution layer which is composed of 32 filters of size 3*3, after this convolution, 32 feature maps of size 28*28 will be created. Then we apply Maxpooling to reduce the size of the image and the amount of parameters and calculation. At the output of this layer, we will have 32 feature maps of size 13*13. We repeat the same thing with the second layer of convolution this layer is composed of 64 filters, the activation function ELU [13] is applied always on each convolution. A Maxpooling layer is applied after the second convolution layer. At the output of this layer, we will have 64 feature maps of size 3*3.

In order not to fall into the problem of overlearning, we must use the dropout instruction, which is very effective for neural networks. It allows us to deactivate a number of neurons according to our configuration, which will also be used at the output of the fully connected layer [14]. The feature vector resulting from the convolutions has a dimension of 1600.

Finally, at the output layer, we added the SVM classifier by replacing the "Softmax" activation function by the linear activation function, which allows computing the probability distribution of the 10 classes (number of classes in the MNIST database).

4.6. Results and Discussion

After 100 iterations, the results in terms of accuracy and error are illustrated in the following.

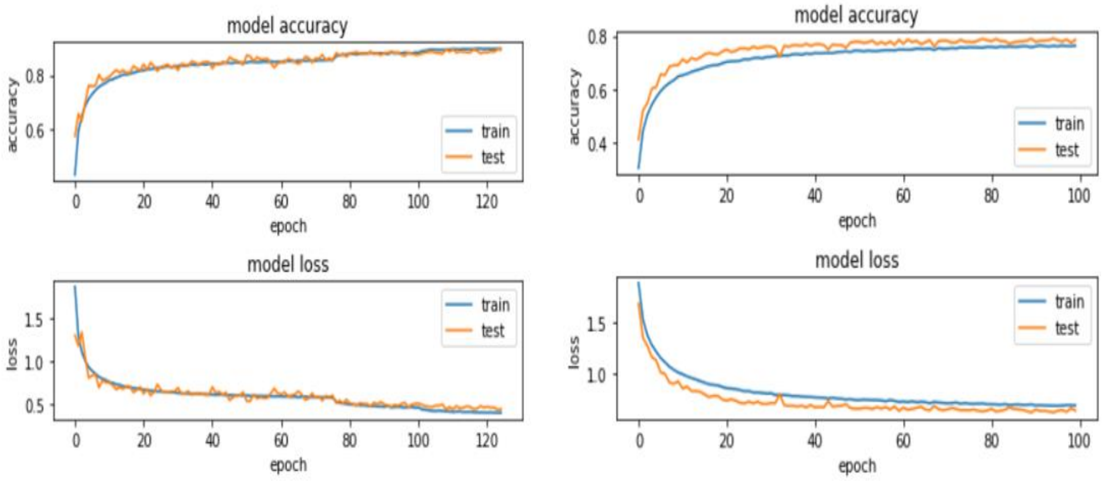

Figure 4. a) Proposed model, b) Based model

From Figure 4, the learning and testing accuracy increases with the number of epochs, this reflects that at each epoch the model learns more information. On the other hand, we find that the learning and validation error decreases with increasing number of epochs.

Table 1. Comparison between the results obtained by the basic model and the proposed model

Model	Accuracy	loss	Execution time	Number of iterations
Based model	98.970%	0.0571	00 :34 :47	100
Proposed model	99.129%	0.0386	00 :50 :21	100

Summarizing the results obtained in the table 1, we notice an error rate of 3.86%, which means that there are 386 images out of 10000 misclassified with an accuracy rate of 99.129%, which largely exceeds that of the basic model.

5. Conclusion

In the end, we developed an implementation for handwritten digit recognition using a classification approach based on convolutional neural network combined with SVM, so we established relationships between error and accuracy via graphs, and for this, we obtained very good results with an accuracy of 99.129% with an error that is almost zero.

The comparison between the results found showed that the increase of the data set, the use of the ELU function, the addition of a Maxpooling layer after the first convolution layer and so on are important players in obtaining better results.

REFERENCES

[1] http://yann.lecun.com/exdb/mnist/

[2] Nedjem. Eddine. Ayat, sélection de modèle automatique des machines à vecteurs de suports : application à la reconnaissance d'images de chiffres manuscrits, thèse doctorat en génie. P. H. D. Université du Québec, Montréal, Canada, 2004.

[3] Matougui. Farida, système interactif pour la reconnaissance des Caractères arabes Manuscrits, SIRCAM, Thèse de Magistère, C, D, T, A, 1990.

[4] Jean-Pierre – "reconnaissance de l'écriture manuscrite", Thèse, Université de Rennes 1.

A. Gattal – "Segmentation automatique pour la reconnaissance numérique des chéques bancaires algériens", Thèse, Thèse de MAGISTER, centre Universitaire de Khanchela, 2011.

[5] https://www.nist.gov/srd/shop/special-database-catalog

[6] www.tensorflow.org

[7] https://keras.io/

[8] C.-L. Huang et C.-J. Wang – "A ga-based feature selection and parameters optimization for support vector machines", Expert Systems with applications 31 (2006), no. 2, p. 231–240.

[9] Guyon, J. Weston, S. Barnhill et V. Vapnik – "Gene selection for cancer classification using support vector machines", Machine learning 46 (2002), no. 1, p. 389–422.

[10] D. Abdelhakim – "La reconnaissance des chiffres manuscrits par les machines à vecteurs de support (svms)", Thèse, Thèse de Master, Université de Tébessa, 2011.

[11] P. Buyssens – "Fusion de différents modes de capture pour la reconnaissance du visage appliquée aux e_transactions", Thèse, Université de Caen, 2011

[12] Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus). arXiv preprint arXiv :1511.07289, 2015

[13] T. C. Faltemier – "Flexible and robust 3d face recognition", Thèse, Citeseer, 2007.