2018 International Conference on Applied Smart Systems (ICASS'2018)
24-25 November 2018, Médéa, ALGERIA

Traffic signs recognition with deep learning

Djebbara Yasmina
Department of Electrical Engineering
Ecole Nationale Supérieure de Technologie
Algiers, Algeria
y_djebbara@enst.dz

Rebai Karima
Department of Electrical Engineering
Ecole Nationale Supérieure de Technologie
Algiers, Algeria
karima.rebai@enst.dz

Azouaoui Ouahiba
Division Productique et Robotique
Centre de Développement des Technologies Avancées
Algiers, Algeria
oazouaoui@cdta.dz

Abstract — **In this paper, a deep learning based road traffic signs recognition method is developed which is very promising in the development of Advanced Driver Assistance Systems (ADAS) and autonomous vehicles. The system architecture is designed to extract main features from images of traffic signs to classify them under different categories. The presented method uses a modified LeNet-5 network to extract a deep representation of traffic signs to perform the recognition. It is constituted of a Convolutional Neural Network (CNN) modified by connecting the output of all convolutional layers to the Multilayer Perceptron (MLP). The training is conducted using the German Traffic Sign Dataset and achieves good results on recognizing traffic signs.**

Keywords—Classification, Recognition, Artificial Neural Network (ANN), Convolutional Neural Network (CNN), Multilayer Perceptron (MLP), Deep learning, Artificial Intelligence, Road signs, Autonomous vehicles.

I. INTRODUCTION

Human factor remains the most common cause of road mortality. Indeed, the potentially dangerous choices made by the driver might be intentional (speed driving, for example) as they might be the result of physical tiredness, drowsiness or a poor perception and interpretation of seen scenes. The introduction of autonomous vehicles will certainly reduce these causes or even make them disappear.

As part of the development of these autonomous vehicles, particularly driving assistance systems, several manufacturers and laboratories have oriented their works towards the exploitation of visual information because of its usefulness for the detection of road, vehicles, pedestrians and traffic signs. The principle of driving assistance systems aiming at road signs recognition is to detect signs, interpret their meaning, then transmit the information to the driver (by a projection on a windshield, a screen or a smartphone) or even better, transmit the information to the vehicle that carries out the execution without needing a human decision. However, given that the classical approach has been bounded by well-structured models of traffic signs (undistorted and completely visible models) only, it became necessary to consider real characteristics of the road environment. For this, the current researches are moving towards the development of recognition systems which are more adapted to real images of road signs that do not generally look like their models.

Motivated by the success of classification and recognition methods, in different domains, based on Deep learning, we are interested in the use of these new advances in Machine learning for traffic signs recognition.

The remainder of this paper is organized as follows; the 2nd section discusses some related works in Traffic Signs Recognition (TSR). In the 3rd section, the datasets used in the development of our approach are presented. Section 4 details the proposed method and section 5 discusses the developed ideas to improve their performances. Section 6 presents the implementation results of the network before and after the application of improvement operations. A summary of the key points and future works concludes the paper in section 6.

II. RELATED WORKS

The last decade shows a growth evolution in the development of intelligent transportation systems (ITS) and especially ADAS and Self-Driving Cars (SDC). In these systems, traffic signs detection and recognition is one of the difficult tasks that confront researchers and developers. This issue is addressed as a problem of detecting, recognizing, and classifying objects (traffic signs) using computer vision and still be a challenge until now.

The work presented in this paper focuses on traffic signs recognition without the consideration of the detection step. For this purpose, this section discusses only related works from this angle. Traffic signs recognition is divided in two parts: features extraction and sings recognition. In the first step, several methods have been proposed, including edge detection [1], scale invariance feature (SIFT) [2], speeded-up robust feature (SURF) [3], Histogram of gradient (HOG) [4] and others. In [5], Bag of Words (BOW) exploiting SURF and k-means classifier was used. Typically, the output of this step is the input of the classification algorithms for the recognition of the road signs. Many algorithms have been used such as K-Nearest Neighbor (KNN) classifier [3], Support Vector Machine (SVM) [6] and neural network [5][7] for traffic signs classification. Authors in [5] proposed the evaluation of three methods namely, Artificial Neural Network (ANN), Support Vector Machine (SVM) and Ensemble subSpace KNN using

978-1-5386-6866-5/18/$31.00 ©2018 IEEE

BoW where every road sign is encoded with 200 features. The Multi-layer Perceptron Neural network provides better results.

Currently, Convolutional networks are gradually replaced traditional computer vision algorithms for different applications such as object classification and pattern recognition [7][8]. It is used for the extraction and the learning of depth description of the traffic signs. This solution overcomes the step of descriptors extraction which is very sensitive to different factors. This network takes 2D image and processes it with convolution operations. It has the ability to learn a representative description of image.

III. TRAFFIC SIGNS DATASET

A rich dataset is needed in object recognition based on neural network in order to train the system and evaluate its results. For the purpose of traffic signs classification, we used the German Traffic Sign Benchmark (GTSB) [9] which contains 43 classes divided into 3 categories as represented in table I.

TABLE I. THE DATASET DISTRIBUTION

Category	**Task**	**Number of images**	**Shape**
Training data	Used to train the network	34799	4 dimensions tensor to determine the image index in the dataset, the pixel's row-column and the information it carries (Red Green Blue value)
Validation data	Allows to supervise the network performances while training it (a reduced version of testing data)	4410	
Testing data	Used to evaluate the final network	12630	

IV. THE NEURAL NETWORK ARCHITECTURE

Using a fully connected neural network to make an image classification requires a large number of layers and neurons in the network, which increases the number of parameters leading the network to over-fitting (memorizing the training data only). The input image may also lose its pixels correlation properties since all neurons (carrying pixels values) are connected to each other [7].

Convolutional neural networks have emerged to solve these problems through their kernel filters to extract main features of the input image and then inject them into a fully connected network to define the class [7].

The chosen architecture in our application is LeNet-5 convolutional neural network (Fig. 1) firstly used for handwritten digits recognition [10]. It contains 9 layers: 5 layers of convolution and simplification functions made by 22 5x5 kernel filters and a max pooling filter of 2x2 to reduce at last the input image of 32x32 into 16 maps of 5x5. The feature images carry most important features to define a specified traffic signs class by processing them into a 4 layers fully connected network.

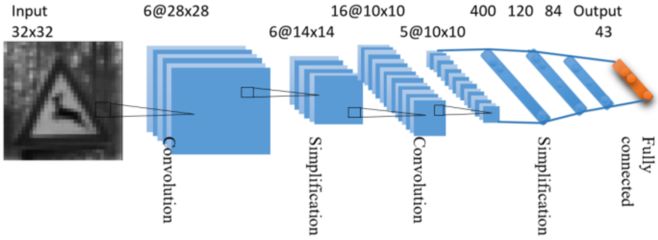


Fig. 1. LeNet-5 architecture

The training phase of our neural network updates its parameters Φ (weights and biases) in order to reach an adequate accuracy value. The update algorithm chosen in our application is a supervised learning algorithm called gradient descent with mini-batches where a multi-dimensional error function *C* (depends on all the network parameters, over 70 000 in the case of LeNet-5) is calculated over mini batches of 64 training examples (to avoid calculus over 34799 images at every stage). Once the error function is obtained, the algorithm will search for the function's decreasing direction by using the gradient of each parameter and then update them under the formula (1) [8], where γ is the learning rate:

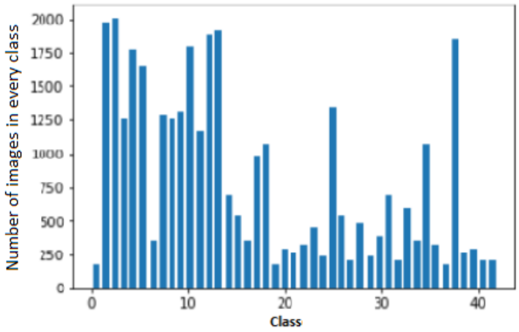
$$\Phi_{t+1} = \Phi_t + \gamma \, \nabla C(\Phi) \tag{1}$$

The algorithm repeats the described process until it reaches the desired results. At the end, the parameters of the neural network are well trained to know what features the network must extract (convolution phase) and which class it must attribute to the input (classification phase).

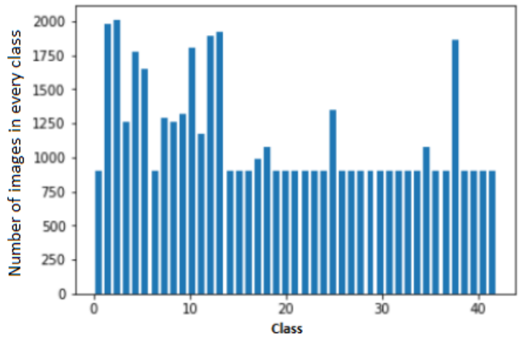
V. PERFORMANCE IMPROVEMENT

A. Training data

The unbalanced distribution of images in the German Traffic Sign Benchmark privileges some classes over others during the training phase because they are better represented in terms of number of images. In order to make sure that the learning of the network is well performed, a data augmentation of some classes is done by applying some geometric transformations (rotation, translation, and shear mapping) on many of their images as shown in Fig. 2.



(a) Distribution before data augmentation



(b) Distribution after data augmentation

Fig. 2. Comparative histograms of data augmentation

The algorithm takes only classes with less than 1000 images to pick randomly images and makes one of the transformation operations (Fig. 3). The resulting images are added to the same class until its elements number reaches the bias which is 1000 images.

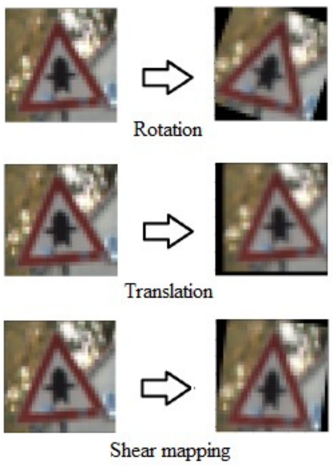


Fig. 3. Geometric transformation for data augmentation

B. The network architecture

In its established architecture, LeNet-5 only takes features resulting from the second convolution operation while the ones of the first convolution might contain elements as important as the ones injected into the fully connected network. Considering this hypothesis, we performed a modification on the first layer of the fully connected network by adding the results of the first convolution operation. This layer will then become a 1576 neurons layer instead of 400 neurons layer as shown in Fig. 4.

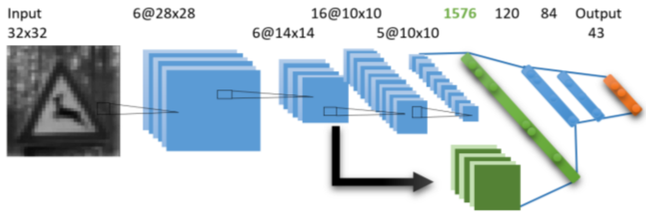


Fig. 4. Modified LeNet 5 architecture

C. The dropout operation

In neural network training, the dropout method is established to prevent from over-fitting by shutting down some neurons during the training phase to give the network a flexible margin to react to inputs out of the training examples data [11]. In our case, we performed a shutting down of 30% and 50% of the total neurons of the second and third layers of the fully connected network respectively. A 90% dropout (shutting down 10% of neurons) on the layer preceding the fully connected network is used in addition to the previous operation.

VI. Implementation Results

To built and train the network, the TensorFlow deep learning library [12] is used. Training and testing were implemented using the dataset described in section III and the developed method succeeds in classifying the 43 traffic signs classes.

The implementation results of the network LeNet-5 and its improvement operations show the impact of each changed element. As represented in the curves of Fig. 5, the enrichment of the first layer of LeNet-5 fully connected network made the validation accuracy jump from 94,7% (blue) to 95,1% (green) after 120 iterations of the learning algorithm. The new given architecture can now combine between many more factors to classify traffic signs. After applying data augmentation (red), an accuracy of 95,3% at the 120th iteration is noticed, making the network performances become even better than the last ones. It is also due to new balanced property of the training data in different classes.

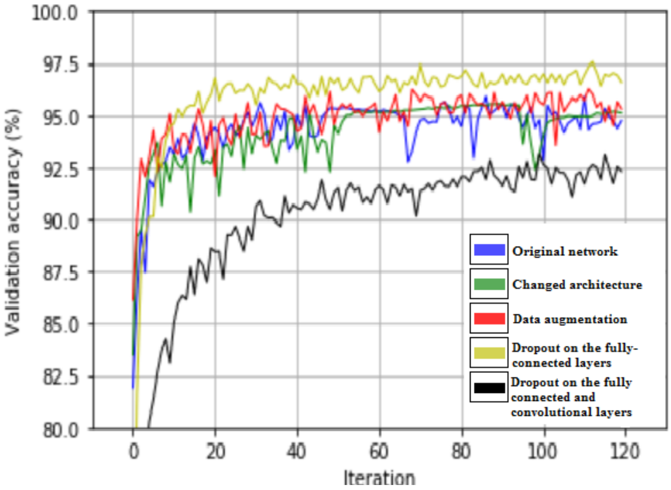


Fig. 5. Accuracies evolution after different implementations

In addition to the augmentation of the learning data, a dropout on 2 layers of the fully connected network is applied and obtained result is a 97,1% validation accuracy (yellow) which corresponds to a 95,2% test accuracy. However, when the dropout is applied to the fully connected network and one of convolutional layers the performances decreased from 97,1% to 92,5% validation accuracy (black). This degradation is explained by the fact that applying a dropout on a convolutional layer means that some neurons which hold convolution results already judged essential to make an object recognition are shut down.

The obtained results show the effectiveness of the developed method since a validation accuracy of 97,1% is achieved. However, the built network could not classify some examples correctly as illustrated in Fig. 6. This figure illustrates, in gray levels, the accuracies of correctly recognized traffic signs and the network confused recognition cases. It is clear that the proposed network achieves 100% recognition rate for some traffic sign classes. The matrix also shows that the most confusion situations (in the red circle) concern classes from 20 to 31 represented in Fig. 7.

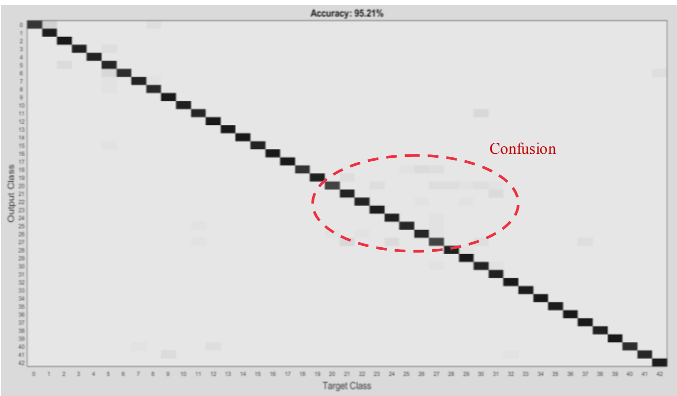


Fig. 6. Confusion matrix

An important confusion has been observed between classes 21, 24 and 27, which is due to their common category (warning signs) and their containing of linear symbols.

Fig. 7. Examples of confusion between traffic signs

VII. CONCLUSION

This paper presented a convolutional neural network implementation used for traffic signs recognition. The basic proposed network together with the different improvement operations allowed us to be aware of which parts and phases that have the control on the system reliability. It is likely that we would obtain better results by reinforcing the convolution stage of the network with more layers in order to extract more features. It also would be interesting to exclude confusions by comparing classes with the highest proportions in the confusion matrix and pull out their common factors to reverse them by image adjustment.

REFERENCES

[1] P. Dewan, R. Vig, N. Shukla and B. K. Das, “An Overview of Traffic Signs Recognition Methods,” International Journal of Computer Applications, Vol. 168 – N..11, June 2017

[2] D. Jianmin and V. Malichenko, “Real time road edges detection and road signs recognition,” IEEE International Conference on Control, Automation and Information Sciences (ICCAIS), Changshu, China, 29-31 Oct. 2015

[3] Y. Han, K. Virupakshappa, E. Vitor, S. Pinto and E. Oruklu, “Hardware/Software Co-Design of a Traffic Sign Recognition System Using Zynq FPGAs,”, In Electronics journal, 2015, Vol. 4, p. 1062-1089; doi:10.3390/electronics4041062.

[4] F. Zaklouta and B. Stanciulescu, “Real-Time Traffic-Sign Recognition Using Tree Classifiers,” IEEE Transactions On Intelligent Transportation Systems, Vol. 13, N. 4, December 2012, p. 1507-1514.

[5] K. Tohidul Islam, R. Gopal Raj and G. Mujtaba, “Recognition of Traffic Sign Based on Bag-of-Words and Artificial Neural Network,” Symmetry journal, 2017, Vol. 9, 138; doi:10.3390/sym9080138.

[6] S. Maldonado-Bascón, S. Lafuente-Arroyo, P. Gil-Jiménez, H. Gómez-Moreno, and F. López-Ferreras, “Road-Sign Detection and Recognition Based on Support Vector Machines,” IEEE Transactions On Intelligent Transportation Systems, Vol. 8, N. 2, June 2007; p. 264-278.

[7] L. Abdi, "Deep learning traffic sign detection, recognition and augmentation," Proceedings of the Symposium on Applied Computing, Maroc, 2017, p. 131-136.

[8] Y. Moualek, "Deep learning pour la classification des images," Master's thesis, Abou Bakr Belkaid University, Tlemcen, 2017.

[9] http://benchmark.ini.rub.de/?section=gtsrb&subsection=dataset

[10] Y. Le Cun, L. Bottou, Y. Bengio and P.Haffner, "Gradient-Based learning applied to document recognition," Proceedings of IEEE, Vol. 86, N°11, p. 2278-2324, 1998.

[11] N. Srivastava, G. Hinton, A. Krizhevsky, I. Stuskever and R. Salakhutdinov, "Dropout : A Simple Way to Prevent Neural Networks from Overfitting," Journal of Machine Learning Research, Vol. 15, 2014, p. 1929-1958.

[12] https://www.tensorflow.org