VOL. 12, NO. 20, OCTOBER 2017 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

CONVOLUTIONAL NETWORK FOR TOOL DISCRIMINATION

Robinson Jiménez M.[1], Oscar Avilés S.[1] and Diana Ovalle M.[2]
[1]Mechatronics Engineering Program, Faculty of Engineering, Militar Nueva Granada University, Bogotá, Colombia
[2]Electronics Engineering Program, Faculty of Engineering, Distrital Fco. José de Caldas University, Bogotá, Colombia
E-Mail: oscar.aviles@unimilitar.edu.co

ABSTRACT

The present paper discusses the use of deep learning techniques, in particular a convolutional neural network, which is trained to identify, in an image, a surgical cutting tool located on a plane. Initially a database is established regarding the tool with different rotations and after this, the base structure of the convolutional network for its training is determined. It is possible to obtain an average identification percentage of 89% with respect to its discrimination in a group of tools also of surgical cut.

Keywords: neural network, identification, convolutional.

INTRODUCTION

Pattern recognition techniques have been supporting the development of artificial intelligence systems for several decades. In turn, for more than half a century neural networks have been one of the techniques of recognition of patterns more worked and with which it is sought to give greater autonomy to robotic agents, computation and the like. Among the most recent techniques of neural networks, deep learning is highlighted, generally based on multilayer neural systems, that seek to emulate the human brain and in which various methods are used for the training of hidden layers, such as gradient decent techniques and restrictive Boltzmann machines (RBM), and others.

The main developments in Deep Learning (DL) cover some specific cases such as modeling of time series data [1] and based on these, applications are presented e.g. solar radiation prediction [2]. Another common application of this technique is speech recognition systems [3] and character recognition systems [4]-[7]. Although applications cover a large number of fields, for example, in [8] a medical prediction of the pathologies of two types of erythema is made, DL techniques are still under development.

For the development of the this paper, it is of interest the applications of DL in relation to the image processing. Regarding to this, it has been begun to present developments of methods of object identification and extraction of characteristics in two dimensional images. For example, in [9] the authors propose a variation to the training of a neural network DNN (Deep neural networks), a type base of DL, with the aim of extracting from complex satellite images desired patterns such as vehicular detection.

Other applications focus on facial detection, for example, in [10] authors use a particular DL technique to identify images, convolutional networks. In this case faces are identified in conditions with changes of pose, expression and illumination, taking a base of 117 thousand faces with these changes in order to recognize 5 poses that with variations of angle determine 15 possible subcategories in which to find a possible face. The convolutional neural networks (CNN) are the basis of the development presented in this paper and their understanding can be reached in greater depth in [11]. Other applications of face detection using DL allow identifying states of numbness in a driver [12], expression recognition [13] and recognitions of facial beauty features [14] [15].

In this work the training of a convolutional network for the identification of surgical type tools is presented, that in the future allows to train a care robot, starting from an image in real time, it is sought to discriminate the desired tool of a group of tools of the same type.

The irregular shapes of the objects to detect and the size of the image, make that Deep Learning offers great advantages in relation to other machine learning techniques, therefore the efficiency of this technique is evaluated in the discrimination of objects of different nature, giving robustness to the identification in relation to the rotation and translation of the tool in the captured image.

The article is organized as follows. Section 2 presents the conditions of deep training. Section 3, an analysis of results. Finally, in section 4, the conclusions obtained.

CNN TRAINING

Deep learning networks (DL) have emerged as a solution to the multi-layer network training problem. The problems in the training of the second hidden layer onwards, where there was no improvement in the learning ratio or did not converge to a value in the weights of the neurons of each layer, have been solved through systems such as restrictive Boltzmann machines (RBM), deep belief networks (DBN) or convolutional neural networks (CCN). CCN in particular have shown a high performance in image recognition.

The convolutional neural networks are in particular a variation of multilayer perceptron neural networks. This type of network is based on a layered cell array, under a bioinspired model in the cortex of the human brain.

The typical input layer of a convolutional neural network corresponds to a group of feature maps, each map

is obtained by the repeated application of a function through sub-regions of the entire input image, i.e. the convolution of the input image is performed with a linear filter, the characteristics of this filter determine the training of the network, based on the object to be recognized. In (1) it is established the mathematical relation for the calculation of weights in the k map of characteristics, of a determined and denoted layer hk, the filters of this map are determined by the weights Wk and the corresponding bias bk, fulfills the function analogous to the training procedure of a perceptron type network.

$$h_{ij}^{k} = \tanh\left((W^{k} * x)_{ij} + b_{k}\right) \quad (1)$$

The convolution for a two-dimensional signal as the image to be treated is described by (2).

$$o[m,n] = f[m,n] * g[m,n] = \sum_{u=-\infty}^{\infty}\sum_{v=-\infty}^{\infty} f[u,v]g[m-u,n-v] \quad (2)$$

Where multiple maps of characteristics are used in the hidden layers with weights Wx, with several orientations. For the case of object identification in this paper, it is opted for the training of a convolutional neural network, whose graphic structure is represented in Figure-1.

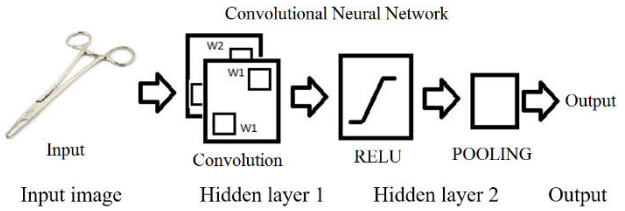

Figure-1. Layer structure of the CNN used.

The presented neural structure has a convolution layer for grayscale images, i.e. one dimension, of a known high (H) and width (W). Figure-2 illustrates the different hyper parameters to be considered in the training, which are described below.

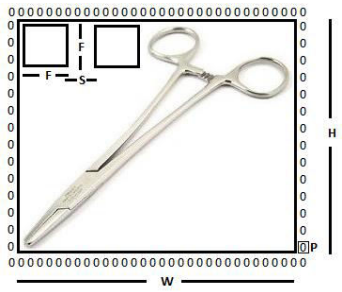

Figure-2. Convolution hyper parameters.

Filter (F): it has high and wide dimensions, for case F but may not be symmetrical, the depth of the filter is D1 and also depends on whether the image is color or grayscale.

Stride (S): range from which to apply the filters to the input volume.

Padding (P): or filled with zeros, it extends the sides with zeros.

So, that the result of the convolution enters a layer called RELU (rectified linear unit), which is an activation function layer. The output volume dimensions of the convolution layer to the RELU are calculated by (3).

$$W_2 = \frac{(W_1 - F + 2P)}{S} + 1$$
$$H_2 = \frac{(H_1 - F + 2P)}{S} + 1 \quad (3)$$
$$D_2 = k$$

The pooling layer operates independently per layer and progressively reduces the size of the layers by maximum or average methods.

As a recognition object, hemostats are used, which are shown in Figure-2, for which it is necessary to establish a wide set of samples, for this case, 100 images are taken in different orientations and location in a support plane, as shown in Figure-3.

Figure-3. Layer structure of the CNN used.

The first step corresponds to the uniformity in the size of each image of the support plane and tool used for training, so all are scaled to a scale of 64×64 pixels, in order to balance the number of samples and the size of each, which directly affects the architecture of the convolutional network and processing times. For this, convolution filters of size 9× 9, stride of 1 and padding in 0. Figures 4 and 5 illustrate the result of the convolution filters.

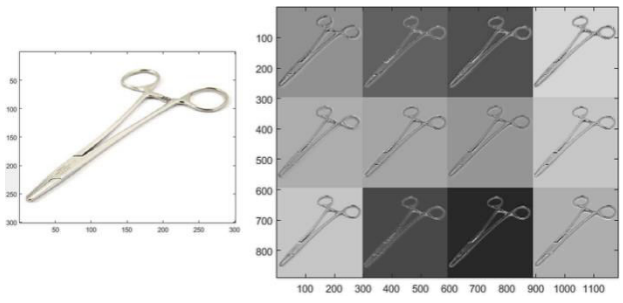

Figure-4. Filter 1 response.

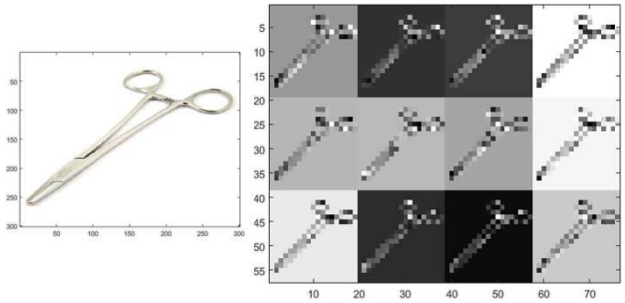

Figure-5. Filter 2 response.

Figure-6 illustrates the result of the RELU layer and Figure-7, the result of the pooling layer for each of the two filters using the averaging method.

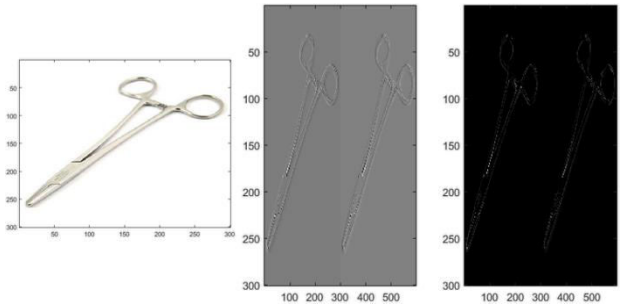

Figure-6. RELU layer response.

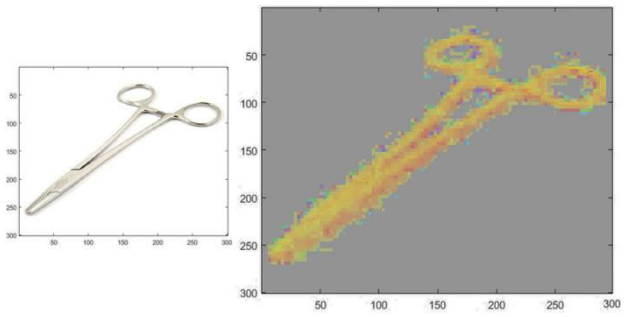

Figure-7. Pooling layer response.

The data matrix obtained at CNN output, which corresponds only to the concentrated information of the pixels in the region of each tool, is entered into a conventional back propagation neural network that is responsible for determining the final classification with respect to two classes, known hemostats or not known.

RESULTS AND DISCUSSIONS

The development was implemented in a 3.2 GHz core i7 computer and 8MB of RAM, under the MATLAB® programming environment, using the tool matconvnet 1.0 beta 16, which allows the implementation of convolutional networks in this environment.

To evaluate the discrimination of the desired tool, a group of unknown surgical tools were presented to the final network. Figure-8 illustrates the four tools for evaluating the training and recognition of hemostats, the last one being tool 1, and in order from left to right according to Figure-8, tool 2, 3, 4 and 5.

Figure-8. Pooling layer response.

The evaluation of the trained network is validated with different orientations of each tool, in order to seek to give independence regarding the position of the object. In the classification phase at the output of the convolutional structure, the neural network generates a response in the range 0-1; this allows determining the approximation of the input object with respect to those known by the trained network. Figure-9 illustrates the flowchart of the procedure followed for the evaluation.

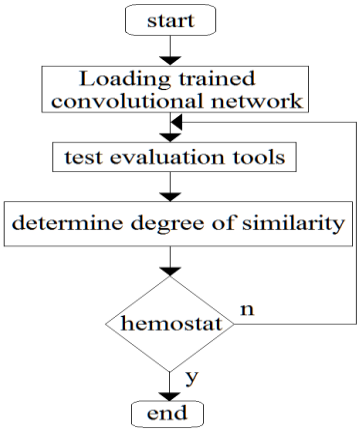

Figure-9. Pooling layer response.

Table-1 illustrates the results obtained in the evaluation of the final trained network. A base of 125 images of the evaluation tools was taken, 25 for each, where the average value of the network output for each tool and its standard deviation is tabulated.

Table-1. Results of the network.

Input image	Number of samples	Average	Standard deviation
Tool1	25	0.89	0.041
Tool2	25	0.16	0.032
Tool3	25	0.84	0.048
Tool4	25	0.13	0.021
Tool5	25	0.81	0.036

It can be seen that the highest values in the output of the network correspond to the tools type clamp, the highest, as expected, would correspond to the desired tool, hemostatic clamps. However, when analyzing the standard deviation, it can be seen that under certain rotational

conditions tool 1 and 3 are confused, while tools 2 and 4 clearly belong to an unknown class.

Finally, given the last case discussed above, only two data groups are presented, the desired and unknown tool, using a dichotomous threshold point derived from Table-1 at 0.85 in relation to the network output. It is validated with images of 100 rotations of the desired tool (number 1) and 100 of the unwanted tool (number 3). The results obtained are shown in Table-2.

It can be seen that in essence the recognition system manages to identify each class with an accuracy degree, related by the matrix of confusion of 82% with respect to the discrimination of each tool.

Table-2. Matrix of confusion of results.

Classification	Hemostat	Not know
True Positive	82	81
False Positive	18	19

CONCLUSIONS

A neural network of the convolutional type was designed and trained for the discrimination of a surgical type tool. Taking as input images of a considerable size (4096 pixels), it has an average response of 620 milliseconds using the equipment described, with a degree of safety of the order of 89%, in the recognition of the tool as a clamp.

The inclusions of tweezing tools generate confusion of classes when subjected to image rotation, reducing discrimination and therefore performance. It is an expected aspect, which would require a confirmation by variation of the camera angle and re-evaluation of the image, to be reduced.

The degree of final discrimination is questioned as to whether it is satisfactory or not, but for an initial test of recognition, a solution is obtained to identify in images subjected to translation and rotation, with an acceptable degree of accuracy.

ACKNOWLEDGEMENTS

The authors would like to thank the Militar Nueva Granada University of Colombia and the Distrital Fco Jose de Caldas University for their support during the development of this project.

REFERENCES

[1] Martin Längkvist, Lars Karlsson, Amy Loutfi. 2014. A review of unsupervised feature learning and deep learning for time-series modelling. Pattern Recognition Letters. 42(1): 11-24, ISSN 0167-8655, http://dx.doi.org/10.1016/j.patrec.2014.01.008.

[2] Ruiz Cardenas Luis Carlos, Jimenez Moreno Robinson, Amaya Dario. 2016. Predicción de radiación solar mediante deep belief Network. Revista Tecnura ISSN: 0123-921X v.20 fasc.47 pp. 39-48.

[3] Xiaodong Cui; Goel, V.; Kingsbury B. 2014. Data Augmentation for deep neural network acoustic modelling. Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. Vol., No., pp. 5582, 5586, doi: 10.1109/ICASSP.2014.6854671.

[4] Walid R.; Lasfar A. 2014. Handwritten digit recognition using sparse deep architectures. Intelligent Systems: Theories and Applications (SITA-14), 2014 9th International Conference on, vol., no., pp. 1, 6, 7-8. doi: 10.1109/SITA.2014.6847284.

[5] Yu Qi; Yueming Wang; Xiaoxiang Zheng; Zhaohui Wu. 2014. Robust feature learning by stacked autoencoder with maximum correntropy criterion. Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on, vol., no., pp. 6716, 6720, 4-9 . doi: 10.1109/ICASSP.2014.6854900

[6] Shusen Zhou, Qingcai Chen, Xiaolong Wang. 2013. Active deep learning method for semi-supervised sentiment classification, Neurocomputing. 120: 536-546, ISSN 0925-2312, http://dx.doi.org/10.1016/j.neucom.2013.04.017.

[7] Wang Y.; Narayanan A; Wang D. On Training Targets for Supervised Speech Separation. Audio, Speech, and Language Processing, IEEE/ACM Transactions on, vol. pp. no. 99, pp. 1, 1doi: 10.1109/TASLP.2014.2352935.

[8] Puerta Felipe, Jimenez Moreno Robinson, Amaya Dario. 2015. Prediction system of erythemas for phototypes i and ii, using deep-learning. Revista Vitae ISSN: 0121-4004 v.22 fasc.3 p. 188-196.

[9] Xueyun Chen; Shiming Xiang; Cheng-Lin Liu; Chun-Hong Pan. 2014. Vehicle Detection in Satellite Images by Hybrid Deep Convolutional Neural Networks. Geoscience and Remote Sensing Letters, IEEE. 11(10): 1797, 1801, doi: 10.1109/LGRS.2014.2309695.

[10] Cha Zhang; Zhengyou Zhang. 2014. Improving multiview face detection with multi-task deep convolutional neural networks. Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on, vol., no., pp. 1036, 1041, 24-26. doi: 10.1109/WACV.2014.6835990.

[11] Zeiler M. D. & Fergus R. 2013. Visualizing and Understanding Convolutional Networks. arXiv preprint arXiv:1311.2901.

[12] Dwivedi K.; Biswaranjan K.; Sethi A. 2014. Drowsy driver detection using representation learning. Advance Computing Conference (IACC), 2014 IEEE International, vol., no., pp. 995, 999, 21-22. doi: 10.1109/IAdCC.2014.6779459.

[13] Inchul Song; Hyun-Jun Kim; Jeon, P.B. 2014. Deep learning for real-time robust facial expression recognition on a smartphone. Consumer Electronics (ICCE), 2014 IEEE International Conference on, vol., no.

[14] Junying Gan, Lichen Li, Yikui Zhai, Yinhua Liu. 2014. Deep self-taught learning for facial beauty prediction, Neurocomputing. 144: 295-303, ISSN 0925-2312, http://dx.doi.org/10.1016/j.neucom.2014.05.028.

[15] Arif Muntasa. 2015. Facial Recognition Using Square Diagonal Matrix Based on Two-Dimensional Linear Discriminant Analysis. International Review on Computers and Software (IRECOS). 10(7) ISSN: 1970-8734. doi.org/10.15866/irecos.v10i7.6623.