Journal of Engineering and Applied Sciences 12 (20): 5323-5331, 2017
ISSN: 1816-949X

Implementation of a Data Augmentation Algorithm Validated by Means of the Accuracy of a Convolutional Neural Network

Paula Catalina Useche M., Javier Pinzon Arenas and Robinson Jimenez Moreno
Facultad de Ingenieria Universidad Militar Nueva Granada, Bogota, Colombia

Abstract: The following study presents the validation of an application developed in MATLAB® for data augmentation which allows to improve the training of convolutional neural networks. The validation is done by comparing the accuracy percentages in the prediction of a trained convolutional neural network with five databases augmented in a different way which allows to determine the characteristics of the training images that produce an increase in network recognition capacities. Each network trained was evaluated by confusion matrices and compared their activations against a test image where it was found that the network with the greatest recognition capacity depends on the changes generated by the data augmentation in the original images (rotations, crops, background changes) as well as the ratio of augmented images and the number of original images used by the data augmentation algorithm developed to produce a new database.

Key words: Deep convolutional neural network, data augmentation, layer activations, confusion matrix, new database, images

INTRODUCTION

The efficiency of the convolutional neural networks is limited by a factor external to the network architecture that influences the recognition of objects in images and of words in audio files and it is the size of the training database which is usually limited and can lead to a reduction in the accuracy and robustness of the network, as discussed by Lemley *et al.* (2017) and Fawzi *et al.* (2016). Facing this problem by Dellana and Roy (2016), data augmentation techniques were used in order to increase the database through transformations in the original images such as the use of filters or noise effects, achieving an increase of about 5% in the accuracy of the trained network.

On the other hand by Shen and Messina (2016), data augmentation techniques were used to generate artificial lines of text with Chinese characters, obtaining a reduction of 10% in the error of recognition against a training with only the original database. In the same way, Zhang *et al.* (2015) and Mclaughlin *et al.* (2015) increased their databases through simple transformations such as rotation, translation and background changes for each image, achieving an increase of about 10% in the accuracy of the convolutional neural network trained.

In view of the benefits presented by data augmentation, several works have been carried out for the development of data enhancement techniques that achieve the highest efficiency in the training of artificial neural networks. For example by Ding *et al.* (2016) three different operations were tested for the expansion of the database of images captured by a Synthetic Aperture Radar (SAR) such as translation, random noise and pose losses, concluding that greater efficiency was achieved with the use of the three operations. In De pontes (Oliveira *et al.*, 2016) intrinsic aspects of training images such as the characteristics of the human face, were considered to improve the estimation of age in people by means of convolutional neural networks.

By this same way by Wong *et al.* (2016), the benefits of the data augmentation applied to three methods of recognition of objects by machine vision were evaluated. The first was by convolutional neural networks trained with backpropagation, the second by convolutional support vector machine and the third by a convolutional classifier by extreme machine learning where it was concluded that a change in the spatial location of the objects in the images improves accuracy and reduces overtraining in a greater proportion than a change in the characteristics of the images. Fawzi *et al.* (2016) in contrast an adaptive algorithm was developed that selects the transformations that must be applied to the original database to increase the accuracy and robustness of a deep neural network, achieving an error reduction of 3% more than with the generation of random transformations on training images. On the other hand by Lemley *et al.* (2017), it was developed an intelligent data augmentation algorithm that consists of a neural network that learns how to increase the database during

Corresponding Author: Paula Catalina Useche M., Facultad de Ingenieria Universidad Militar Nueva Granada, Bogota, Colombia

the training process of a convolutional neural network where it is possible to minimize the error and losses of the network for different databases such as faces and places of the world.

Faced with recognition objectives other than the classification of elements in images (Cui *et al*., 2015) and (Salamon and Bello, 2017) investigated the effect of applying data augmentation on the recognition of sounds and words, finding improvements in the performance of convolutional neural networks and deep neural networks by generating an increase in Vocal Ttract Length Perturbations (VTLP) and in Stochastic Featured Mapping (SFM) of the original database.

The following paper presents the development of a Data Augmentation algorithm elaborated in MATLAB® that allows to generate more than 18,000 images by category from a database of 50 white background images of each category and 30 random images. The algorithm provides an innovative method for performing random crops on training images in addition to a new technique of repositioning elements that allows to vary the locations of a tool in a general image. Likewise, a conclusion is reached on the relevant aspects that an increased database must fulfill in order to increase the robustness of the network instead of hindering its learning.

On the original database, rotations, random crops, resizing, changes of position and variation of backgrounds are made in order to obtain large databases with different characteristics from a general database. With each database a convolutional neural network (Krizhevsky *et al*., 2012) was trained to evaluate the variation in the percentage of recognition accuracy of said network which is quantified by means of confusion matrices in order to compare the results obtained with the purpose of determining the characteristics of the databases that generate the highest accuracy.

MATERIALS AND METHODS

Algorithm: The algorithm developed for the generation of data augmentation starts from a test database, consisting of three different tools: surgical scalpel, scissors and screwdrivers.

As the first aspect of increasing the database, a random background change is made to each tool category. Then, each tool is rotated and another background change is generated to obtain a new image, then the rotated tool is taken and, if the space occupied by the image is less than a specified limit, a random crop is made around the tool or on it. Finally, the rotated image is taken, resized, placed in a different position of the image and a new background added.

In general, for each rotation, 3 different images are saved: one of the image of the rotated element, another of the same element rotated plus a random crop and a new background and another of the same element rotated but located in another section of the image and with another background, therefore, if the image is rotated 10 times, at each rotation angle 3 images are saved. In addition, the original image is saved once with background, so that, 31 images are generated in the database increased from a single image of the original database.

In order to calculate the maximum or minimum quantity of output images that can be obtained with a single input image according to the degree of rotation (r) and the final inclination of the image (R), Eq. 1 is used and by means of Eq. 2 the minimum rotation value can be defined to generate a desired quantity of output Eq. 1 for each input image:

$$I = \left(\frac{nR}{r}\right) + 1 \tag{1}$$

$$r = \frac{nR}{I-1} \tag{2}$$

where, n a constant where n = 3 to calculate the maximum value of output images that is generated when at every rotation without exception a random cut is made on the rotated image. n = 2 to calculate the minimum amount of images generated when no cuts occur on any image or n = 1, when the algorithm is partially used, i.e., only rotations of the object are generated without cuts or changes of position. In case of only generating background and position changes, Eq. 3 is used where N is the number of original images:

$$I = 2N + 1 \tag{3}$$

Initialization of algorithm: The algorithm performs an initial reading of all images, both training and backgrounds to know the total amount of images to be processed. It allows the user to enter R and r values to specify the rotation of each object, define the dimensions of the output images (M×M) and specify the minimum amount of images per category to generate the augmented database.

Background change: The program selects a random image from each category of the original database and a random image from the backgrounds folder. It evaluates pixel by pixel of the original image and replaces all the white pixels of that image by the pixels of the selected background so that the background is added to the resulting image as shown in Fig. 1.

The algorithm avoids repeating the already used images of both the database and the backgrounds by means of an array with random numbers of the size there of without being repeated in order to assure the greatest

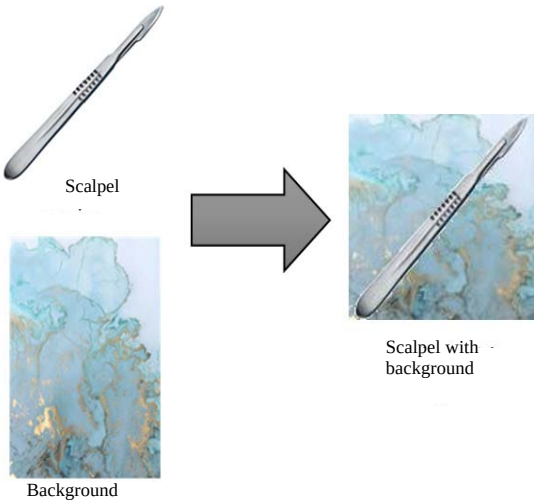

Fig. 1: Background addition

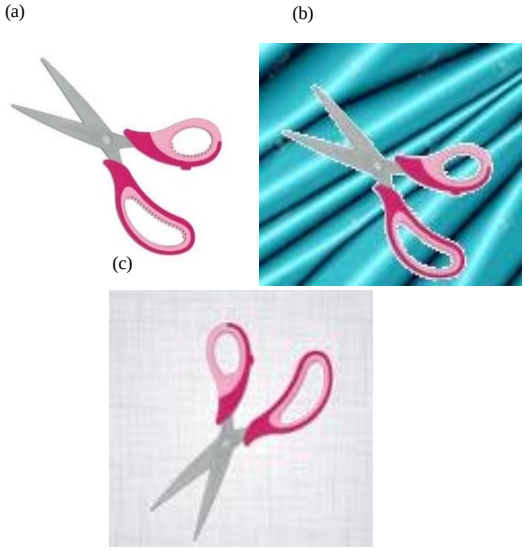

Fig. 2: a) Original image; b) image with background without rotation and c) image rotated with new background

possible diversity among the images of the augmented database. Once all the images are used, the array is restarted with new random numbers.

Algorithm rotation of image: After saving the image with background, the original image is taken again is rotated a value in degrees and to that new image is added another random background as shown in Fig. 2a-c where an image of the "scissors" category (Fig. 2a) was selected, rotated 90° and assigned a new background.

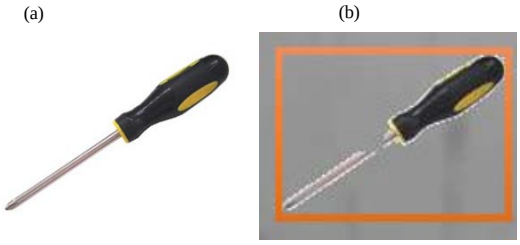

Fig. 3: a) Original image and b) image with cropping on the tool

Random crop on image: The objective of cutting fragments of the image is to incite the network to learn not only the object in its entirety but also parts of it, so it is sought to generate square cuts of random dimension and position that eliminate or cover part of the tool as shown in Fig. 3a, b where a box equal to the average background color, covers part of the screwdriver, separating the tool into two parts.

In order to increase the probability that the cut is made on the tool and not on the background, only those objects that do not occupy much space in the image are selected, i.e., that by covering them with a box as in Fig. 3b, the difference between the dimensions of said box and the dimensions of the total image are less than a threshold value, chosen by the user. If the threshold is exceeded no cuts are made.

Random position of the tool: To change the position of the tool in the output image a White Background (WB) of size is created in which the object is to be repositioned and the image of the tool is scaled by random value of between 1.3 and 1.7 in order to reduce it sufficiently so that it can be located in different positions of the image WB and having the size necessary to be differentiated from the background. By means of Eq. 4 the dimension of the scaled tool is calculated which for an input image of 128×128 pixels generates resized images between 98×98 pixels and 75×75 pixels:

$$D_r = \frac{D_i}{B} \quad (4)$$

Where:

D_r = The dimensions of the output image
D_i = The dimensions of the input image

A random point of the WB image is selected to position the resized tool and the white background is replaced by a color one which can be transposed inverted or alternated its pixels to change its location in the output image and generate a new background such as the one

shown in Fig. 4a, b. Changes that are applied to the original background (transpose invert or alternated pixels) are randomly selected.

Graphic user interface: Figure 5 shows the graphical interface used to augment the original database by applying the combination of some of the previously mentioned transformations. The interface allows to enter the folder's path containing the original database, the backgrounds and the augmented database (Sections A, B and C, respectively) as well as entering the final amount of images for each category (Section K) the size of the generated images (Section I) and the transformations to be made on the original database (Section L).

Section D specifies which folder in the original database, train or test is to be augmented and Section E selects the category entry mode; manual mode to choose the number of categories to apply the algorithm (Section F) and write each of their names in table H or the Automatic mode that allows the program to read and expand all categories contained in the original database within the folder selected in Section D.

In Section J the user determines how many degrees the tool will be rotated (Degrees) and the total value of rotation. In the N Section, the results display of Section M is activated or deactivated, showing some of the images generated by the program.

With the button of the Section G, all the folders' paths to be used in the program and its categories are entered and with the button of the Section O initiates the augmentation of the database. The button of the Section P works as an emergency stop.

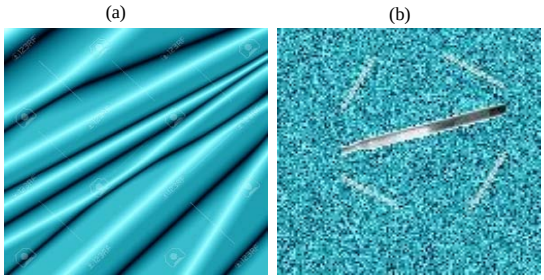

Fig. 4: a) Original background and b) background with random pixels

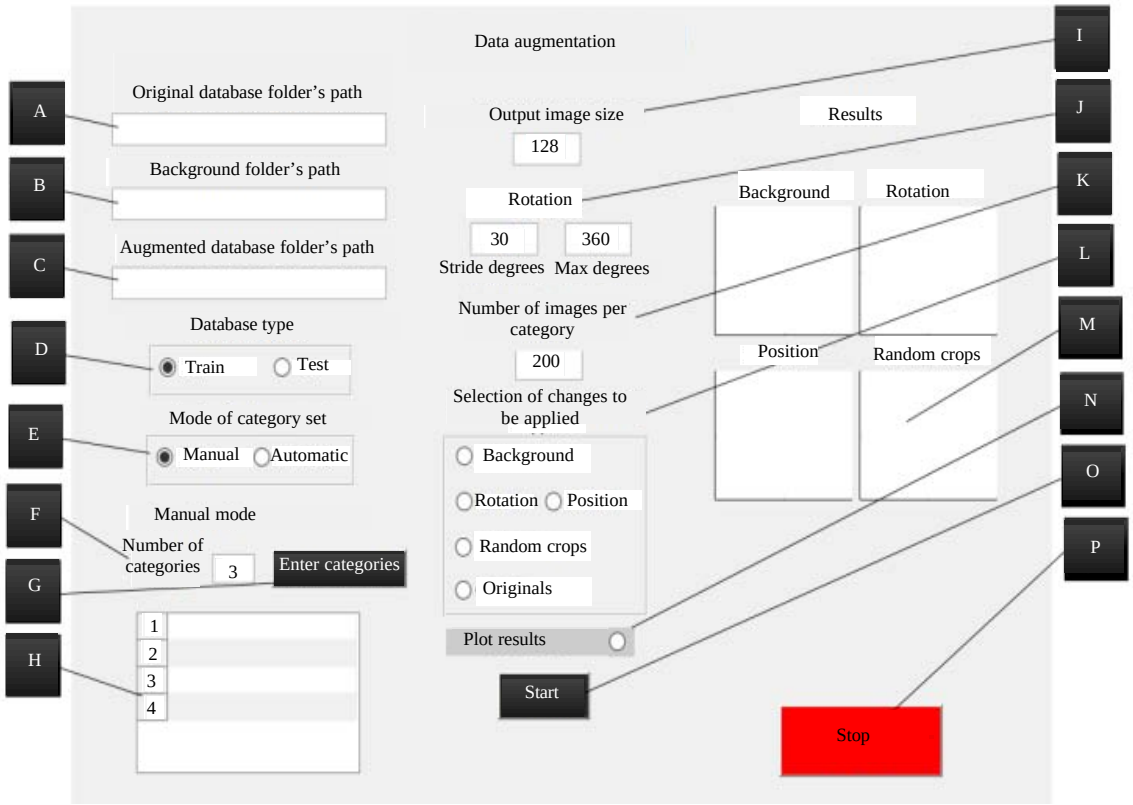

Fig. 5: Graphical user interface of the data augmentation algorithm

RESULTS AND DISCUSSION

From a database of surgical scalpel, scissors and screwdriver with 50 images per category and all on white background, the algorithm developed for data augmentation was applied, generating five augmented databases, each with different characteristics and a maximum of 200 images per category, to obtain a total of 600 training images among the sum of the three categories of tools. The convolutional neural network used has the architecture shown in Fig. 6 for each of the tests where F is the size of the convolution filter, K the number of filters, S the filter step, P the padding, Conv represents the convolution layer, Relu the Rectified Linear Units Layer, Maxpool the Maxpooling layer, Drop the Dropout layer and 50% of the Dropout layer is the disconnection percentage. The input images to the network are in color (RGB channels) and dimension 128×128 pixels.

Augmented database 1 takes random tools from the original database and applies all the changes in Table 1 (background changes, rotations, random cuts and position changes). For the following database the same changes are applied as the database 1 but without generating random crops. In the next two, only changes of background and rotations are generated with variations in the angle of rotation and in the latter the original database is taken and the tools are changed with rotations every 90° without adding any background.

The convolutional neural network of Fig. 6 was trained with each of the databases in Table 1 and tested with a test database of 93 images per category with random backgrounds, rotations and position changes for each tool from which the confusion matrices were generated for each database. Table 2 shows the accuracy results for the network trained with each of the databases (Accuracy), the time required to generate the augmented database (Database time), the epoch training time (Training time) and the number of original images that were used to generate each database (Original images).

Table 1: Augmented databases

Augmented database	Color background	Rotation	Randcrop	Position changes
1	Yes	30°	Yes	Yes
2	Yes	30°	No	Yes
3	Yes	30°	No	No
4	Yes	10°	No	No
5	No	90°	No	Yes

Table 2: Results for each database

Augmented database	Accuracy (%)	Database time (sec)	Training time (sec)	Original images
1	58.78	64.88	60	6
2	65.95	53.59	60	7
3	70.61	73.48	60	19
4	50.18	70.27	63	5
5	53.05	60.12	60	23

The network that obtained the highest accuracy was the one that was trained with the third database where only rotations were made every 30° and random background changes, nevertheless when training the network with those same characteristics but a lower degree of rotation as the fourth database, accuracy

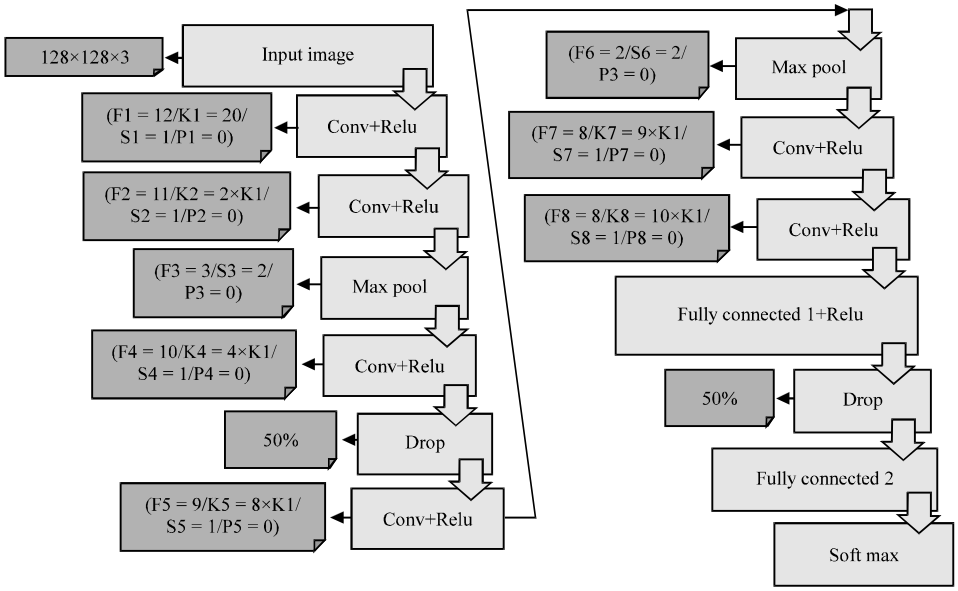

Fig. 6: Architecture of the convolutional neural network

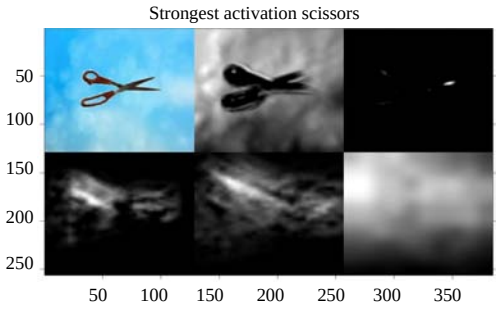

Fig. 7: Stronger activations of the network trained with the databas, Strongest activations scissors

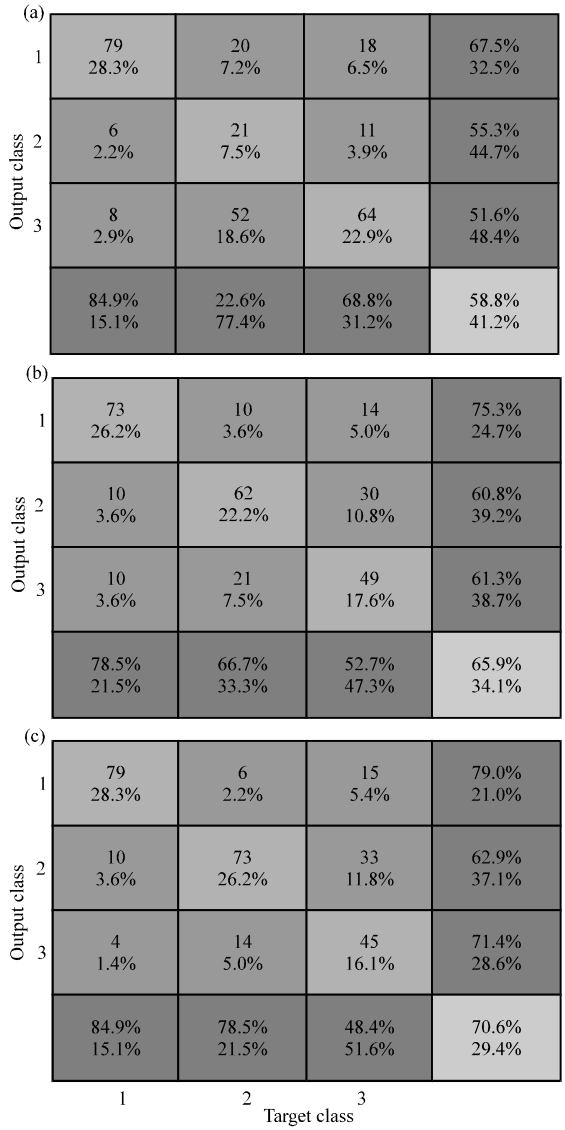

Fig. 8: Confusion matrices for the databases a) 1, b) 2 and c) 3

was reduced by about 20%. The change in accuracy between networks 3 and 4 is because the increase in rotation allows the network to learn the same object in different orientations but it generates more repeated images of the same tool which reduces considerably the number of different objects to learn and leads to memorization rather than the generalization of features.

In addition, it is possible to observe that the database 5, all with white background, obtained a 3% more accuracy than the fourth database due to the number of different images with which the network was trained. However, the lack of backgrounds led it to have the second lowest accuracy, since its stronger activations which indicate the characteristics that takes of each image to perform the recognition, were on the background as can be seen in Fig. 7 where the lighter tones represent a greater activation than the dark ones.

The strongest activation of Fig. 7 in the first relu layer was the background (upper central image) in the second relu layer there was almost no activation (upper right image) and in the following three relu layers, large loss of information was generated (lower images ordered from left to right) where it is not possible to distinguish any part of the geometry of the object.

On the other hand, the network trained with the first database only improved 5% with respect to the fifth database while the second one achieved a 65% accuracy, almost with the same characteristics of the first database but without random crop. To better detail the differences between the three databases with better accuracy, their matrices of confusion and their stronger activations are compared.

Figure 8 shows the confusion matrices for each network where first category corresponds to the scalpel, the second to the scissors and the third to the screwdriver. The “Target Class” columns represent the actual categories of the tools; “Output Class” rows are the classifications made by the network for each category; the percentages of the last row represent the percentage of recognition achieved in each category that is how many images of the same category were adequately classified; the percentages in the right column represent the percentages of successful recognition, i.e., how many elements of a category with respect to the others recognized correctly; and the bottom right corner shows the overall accuracy percentage of the network.

The third database achieved percentages of accuracy by category higher than the other two networks, however in all cases there was a deficiency in the recognition of one of the categories; the screwdriver for the database 3 with a recognition rate of 48.4%, screwdriver for database 2 with 52.7% and scissors for database 1 with 22.6%.

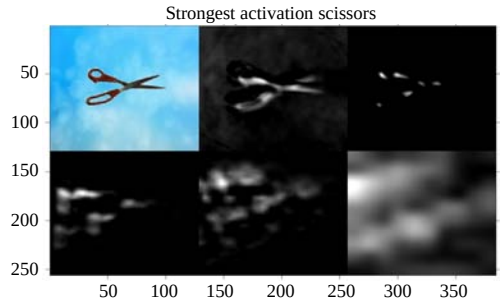

Fig. 9: Activations of the network trained with the database 3, Strongest activations scissors

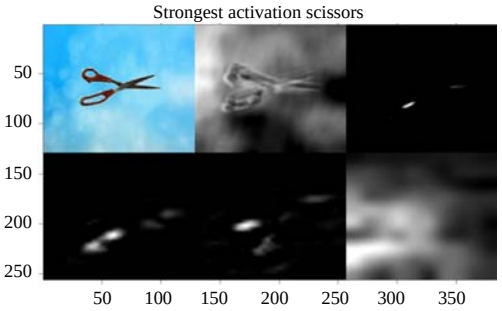

Fig. 10: Activations of the trained network with the database 2, Strongest activations scissors

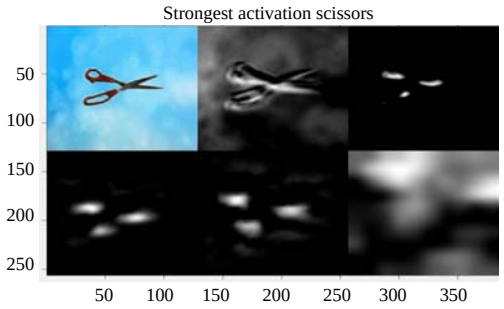

Fig. 11: Activations of the network trained with the first database, Strongest activations scissors

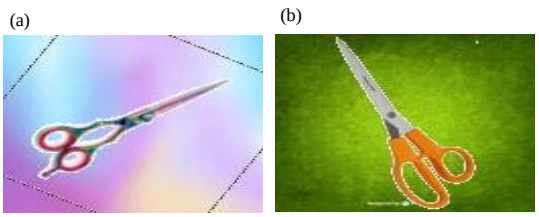

Fig. 12: Differences between; a) training scissors and b) test scissors

In the case of database 3, the activations presented in Fig. 9 focused on the geometry of the tool, however, partial loss of information started from the second relu layer and background activations in the later layers.

For database 2, the greatest activation that was generated in the first relu layer was on the background and the tool as shown in Fig. 10 which means that it failed to recognize the object in order to generate a correct classification but captured the entire image and additionally in the second, third and fourth convolution layers, almost no activation occurred.

On the other hand, the first database presents its strongest activations at the edges of the tool as seen in the first four relu layers and recognizes part of the background in the first convolution layer but less strong respect to the recognition of the scissors as shown in Fig. 11.

The network trained with database 1 presented a greater capacity to recognize the geometry of the element than the other two networks because it managed to capture both the eye rings and part of the tips of the scissors during all the convolutions, however, this had the third best accuracy during the tests. The reduction in accuracy for this network is attributed to the low number of different images with which the augmented database was generated, since it used only 6 of the 50 images of the original database.

On the other hand, unlike the fifth database that had 25 different images but all with the same background, their activations focused on the tool and not on the background which means that it really tries to recognize the element. This change in recognition between the two databases is due to the fact that in addition to requiring a database with different types of tools, it is necessary to vary common aspects of training images such as backgrounds and orientations to prevent the network from memorizing the similar characteristics and not related to the desired learning object.

Due to the lack of diversity in the training tools, the trained network with the first database presents more difficulties recognizing the scissors than the other two tools, since this particular element has geometries and colors more different than the objects of the categories of scalpel and screwdriver as shown in Fig. 12 where one of the training scissors (Fig. 12a) and one of the test scissors (Fig. 12b) are shown with variations in the eye rings and proportions of the tool. On the other hand, scalpels are very similar to each other such as those in Fig. 1 and Fig. 4 and they presented a similar percentage of recognition in all networks.

CONCLUSION

The background changes and rotation of the original images allow to increase more than twice the databases, however when the object to be trained presents different geometries and colors, it is important to maintain a balance between the number of images generated with a single original image and the total amount of original images used, since an imbalance in this relation affects the capacity of recognition of a convolutional neural network with changes of accuracy of about 20% as can be observed with databases 1 and 3.

The database with more diversity of tools for each category and all with different backgrounds achieved the best accuracy however, activations for the network trained with database 1 showed greater learning capacity and better background discrimination during all its convolutions than the other networks, due to the variety of images of the same tool with which it was trained. These results imply that a larger database obtained with the characteristics of database 1 can get to learn the features of the training objects better than with the other databases but requires more than 200 images per category to achieve generalization of each element rather than memorizing.

A possible solution for the unbalance of database 1 is to generate a larger database to encompass many more images from the original database and to generate a greater variety of elements than copies of each one or change parameters of the algorithm of data augmentation to generate fewer random crops, fewer rotations of the same tool or fewer position changes for each object in such a way that the increased database is smaller than with the current algorithm but with more diversity in training images as happened with database 3 which generated about 11 images per element and used 19 different tools per category.

ACKNOWLEDGEMENTS

The researchers are grateful to the Nueva Granada Military University which through its Vice chancellor for research, finances the present project with code IMP-ING-2290 and titled “Prototype of robot assistance for surgery” from which the present research is derived.

REFERENCES

Cui, X., V. Goel and B. Kingsbury, 2015. Data augmentation for deep neural network acoustic modeling. IEEE. ACM. Trans. Audio Speech Lang. Process. TASLP., 23: 1469-1477.

Dellana, R. and K. Roy, 2016. Data augmentation in CNN-based periocular authentication. Proceedings of the International Conference on Information Communication and Management (ICICM), October 29-31, 2016, IEEE, Hatfield, UK. isBN: 978-1-5090-3495-6, pp: 141-145.

Ding, J., B. Chen, H. Liu and M. Huang, 2016. Convolutional neural network with data augmentation for SAR target recognition. IEEE. Geosci. Remote Sens. Lett., 13: 364-368.

Fawzi, A., H. Samulowitz, D. Turaga and P. Frossard, 2016. Adaptive data augmentation for image classification. Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), September 25-28, 2016, IEEE, Phoenix, Arizona isBN:978-1-4673-9961-6, pp: 3688-3692.

Krizhevsky, A., I. Sutskever and G.E. Hinton, 2012. Image net classification with deep convolutional neural networks. Proc. Neural Inf. Process. Syst., 1: 1097-1105.

Lemley, J., S. Bazrafkan and P. Corcoran, 2017. Smart augmentation-learning an optimal data augmentation strategy. IEEE. Access, 5: 5858-5869.

McLaughlin, N., D.J.M. Rincon and P. Miller, 2015. Data-augmentation for reducing dataset bias in person re-identification. Proceedings of the 2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), August 25-28, 2015, IEEE, Karlsruhe, Germany isBN: 978-1-4673-7632-7, pp: 1-6.

Oliveira, D.P.I., J.L.P. Medeiros and D.V.F. Sousa, 2016. A data augmentation methodology to improve age estimation using convolutional neural networks. Proceedings of the 2016 29th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), October 4-7, 2016, IEEE, Sao Paulo, Brazil isBN:978-1-5090-3568-7, pp: 88-95.

Salamon, J. and J.P. Bello, 2017. Deep convolutional neural networks and data augmentation for environmental sound classification. IEEE. Signal Process. Lett., 24: 279-283.

Shen, X. and R. Messina, 2016. A method of synthesizing handwritten Chinese images for data augmentation. Proceedings of the 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), October 23-26, 2016, IEEE, Shenzhen, China isBN:978-1-5090-0981-7, pp: 114-119.

Wong, S.C., A. Gatt, V. Stamatescu and M.D. McDonnell, 2016. Understanding data augmentation for classification: When to warp?. Proceedings of the 2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA), November 30-December 2, 2016, IEEE, Gold Coast, Queensland, Australia isBN:978-1-5090-2896-2, pp: 1-6.

Zhang, C., P. Zhou, C. Li and L. Liu, 2015. A convolutional neural network for leaves recognition using data augmentation. Proceedings of the 2015 IEEE International Conference on Computer and Information Technology: Ubiquitous Computing and Communications: Dependable, Autonomic and Secure Computing: Pervasive Intelligence and Computing (CIT/IUCC/DASC/PICOM), October 26-28, 2015, IEEE, Liverpool, UK. isBN: 978-1-5090-0154-5, pp: 2143-2150.