Contemporary Engineering Sciences, Vol. 10, 2017, no. 29, 1419 - 1429
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ces.2017.710155

RGB-D Training for Convolutional Neural Network with Final Fuzzy Layer for Depth Weighting

Robinson Jiménez Moreno, Oscar Avilés Sanchez

Nueva Granada Military University
Faculty of Engineering. Bogotá, Colombia

Diana M. Ovalle

District University Francisco José de Caldas
Faculty of Engineering. Bogotá, Colombia

Abstract

This paper presents training of novel hybrid network based on three deep convolutional neural network architectures applied to object recognition, based on the depth information supplied for a RGBD camera. For this case, the depth information allows to set the dataset of training images of each network, its architecture and its characteristics, generating a dynamic recognition application by variation of the image capture point, whose final layer is determined by a diffuse inference system. The general architecture designed allows an efficient object recognition applicable to robotic mobile agents, whose perspective of the object varies when approaching or moving away from them, showing an overall performance of 90.19%.

Keywords: Convolutional Neural Network, Robotic applications, 3D environment, fuzzy layer, hybrid network

1 Introduction

In recent years, deep learning techniques have led the field of artificial intelli-

gence [1] where its ability to recognize patterns gives rise to developments such as control systems for vehicular traffic [2]. However, while deep learning improves the capabilities of conventional neural networks, other artificial intelligence techniques such as fuzzy logic offer solutions that do not yet cover these techniques.

The applications of fuzzy systems today are of great relevance in the control of robotic agents, for example, in displacement control applications [3] [4], learning reinforcement for navigation [5] and even for collaborative work among robotic agents [6]. But a robot to navigate and interact with its environment, requires a machine vision system to identify patterns as objects in the scene. In relation to this requirement, specific techniques of deep learning, such as convolutional neural networks (CNN), allow the implementation of efficient image-based artificial intelligence systems [7].

Thus, one of the current fields of application of the CNN is the development of robotic agents based on artificial intelligence, for example, for tasks of recognition of places that can lead to autonomous navigation by the robot [8]. There are also applications for the development of robotic activities in three-dimensional tasks. For example, in [9] and [10], convolutional networks are presented for gripping tasks using a robotic gripper, under aspects such as the depth of the object.

In general, many of the capabilities of neural networks have been potentialized with hybrid networks based on diffuse layers [11]. These diffuse-neural architectures have improved developments such as video sign language recognition [12], voice activity detection [13], graphical data representation [14] and even in robotic applications for obstacle avoidance [15] and path planning [16]. Where it is also possible to find applications of deep learning hybrid networks and fuzzy systems for data classification [17].

This paper presents a novel proposal for the dynamic recognition through convolutional neural networks, in cases where there is a variation of the capture point of the image, i.e., applications where an RGB-D sensor is used that is in the robotic agent and this one moves in relation to an object of interest. This raises the problem of changing the perspective of the object to be recognized, thus proposing, through the work presented here, a solution not found in the state of the art, based on a hybrid diffuse-convolutional network. As mentioned, hybrid models of this type are already presented, such as those presented in [18] and [19], the first case presents the diffuse system as input to the convolutional network while the second is at the output of the same, similar to the case discussed here. However, the problem, the application and final architecture are clearly differentiable.
This paper is divided into 4 parts, where section 2 shows CNN architecture proposed, in others words, describes the layers implemented in this work, convolutional and fuzzy. In section 3 the results are obtained by means of tests in one specific example. Finally, section 4 concludes with respect to the depth recognition algorithm results.

2 CNN Architecture

When a robotic system uses a camera for its interaction with the workspace in which it develops, whether it is a mobile autonomous robot or a robotic arm with a camera in the end effector, the perspective of the object towards or away from it changes. From this point of view, the training of a pattern recognition system, conventional neural or convolutional type, will have different input data for its classification. For example, Figure. 1 presents the resolution changes presented by the approach of an object, which makes the network response vary in the percentage of success and increases the probability of class confusion.

The solution proposed consists of the neural training of a convolutional network specialized in the different perspectives of the object to be recognized, depending on the depth in which it is, i.e., the training of a set of neural networks is done, where each one will learn to recognize the object from the distance and will weigh each network as it approaches the object.

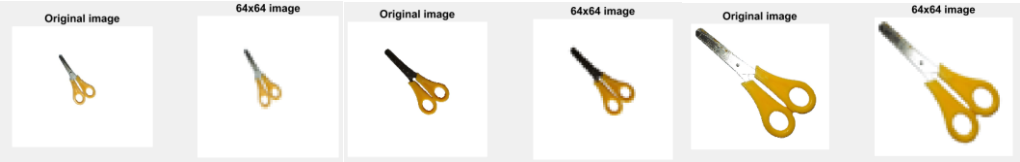

Figure 1. Variation of characteristics when approaching an object.

The architecture of the hybrid CNN that is proposed is illustrated in Figure. 2. Where the input image corresponds to a group of tools in 4 categories: nippers, screwdriver, scalpel and scissors, all in color and taken with a Creative Blaster Senz3D RGB-D camera, whose 3D vision range is from 0.2 to 1.5 meters.

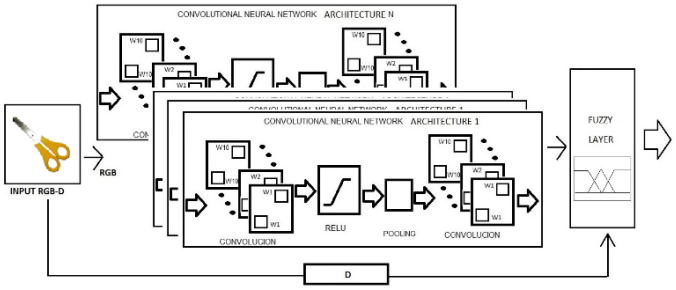

Figure 2. CNN architecture proposed.

The architecture continues with a group of three convolutional neural networks working in parallel, each one trained with databases of 200 images of each category, taken at different distances: 20 cm, 40 cm and 60 cm, correspondingly. The response of each network, by category, is entered into a final layer of diffuse weighting that is in charge of determining the unique output of the network.

Table 1 shows the CNN architecture used for each of the three networks, where square filters are used whose sides are shown in column two, with 70, 90 and 120 filters per convolution layer [20].

Table 1. Architecture CNN

Layer	Kernel	
Input	64x64x3	
Convolution/RELU	5	#70
MaxPooling	4	
Convolution/RELU	3	#90
MaxPooling	2	
Convolution/RELU	2	#120
MaxPooling	2	
Fully-Connected	4	
Softmax	4	

The final weighting layer corresponds to the diffuse inference system illustrated in Figure. 3. The four base inputs of the system are the current depth value of the capture of the image to be classified and the fuzzification of each of the outputs of the 3 CNNs.

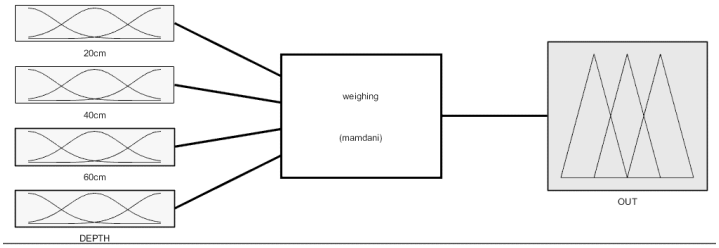

Figure 3. Fuzzy layer. Source: The authors.

Since the output of the CNN classification stage is based on a nonlinear function, Gaussian membership functions are set for each fuzzy set, according to its category, as shown in Figure. 4. In order to uniformly distribute the universe of discourse with intersections of 0.5, which allow a total membership degree of 1 for any input, the output of the diffuse system obeys the same behavior.

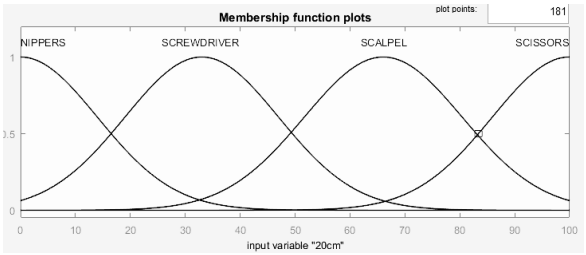

Figure 4. Input CCN fuzzy set.

Figure. 5 shows the fuzzification of the depth level, this distribution is based on the range of vision of the RGB-D camera used and associated to the levels of training depth of each network.

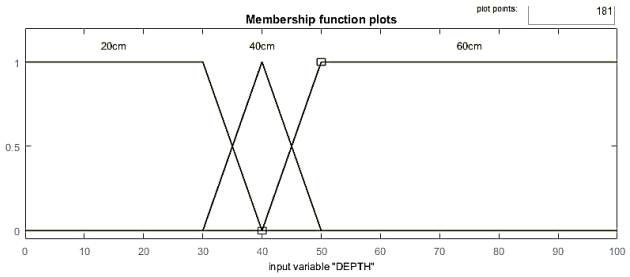

Figure 5. Input Depth fuzzy set.

The fuzzification of each output of the network, i.e., of each category (S_Ctg), should be framed in the same universe of discourse, which is characterized by ease in the range of 0 to 100. For this, it is developed the equations by category that are associated in Table 2, according to the behavior of the membership functions given in Figure. 4.

Table 2. CNN Architecture

CATEGORY	FUZZIFICATION	RANGE
NIPPERS	(1- S_Ctg1)*33.5	0-33.5
SCREWDRIVER	S_Ctg2*33.5	0-67
SCALPEL	S_Ctg3*33.5+33.5	33.5-100
SCISSORS	S_Ctg4*33.5+67	67-100

For the fuzzy relational computation that establishes the output of this last layer of the proposed network architecture, it is used the Mamdani inference algorithm shown in Figure. 6.

1. Compute the degree of fulfillment for each rule by: $\beta_i = \max_X [\mu_{A'}(\mathbf{x}) \wedge \mu_{A_i}(\mathbf{x})]$, $1 \leq i \leq K$. Note that for a singleton set ($\mu_{A'}(\mathbf{x}) = 1$ for $\mathbf{x} = \mathbf{x}_0$ and $\mu_{A'}(\mathbf{x}) = 0$ otherwise) the equation for β_i simplifies to $\beta_i = \mu_{A_i}(\mathbf{x}_0)$.
2. Derive the output fuzzy sets B_i': $\mu_{B_i'}(\mathbf{y}) = \beta_i \wedge \mu_{B_i}(\mathbf{y}), \quad \mathbf{y} \in Y, \quad 1 \leq i \leq K$.
3. Aggregate the output fuzzy sets B_i': $\mu_{B'}(\mathbf{y}) = \max_{1 \leq i \leq K} \mu_{B_i'}(\mathbf{y}), \quad \mathbf{y} \in Y$.

Figure 6. Mamdani algorithm [21]

3 Experimental Results

To evaluate the performance of the designed fuzzy layer, an example of validation is set, where the input of the designed hybrid architecture is entered the image

observed in Figure. 7, whose prediction is on the right side. A representative case is chosen where one of the networks gives a category different from the one entered.

Figure 7. Network prediction

The inputs to the fuzzy system are determined by the equations in Table II, applied to the output of each of the 3 CNNs, which give the prediction values of each category relative to the input image. Table 3 illustrates the results of the fuzzification.

Table 3. Fuzzification inputs

		Fuzzification	
Network	**Category**	**u(x)**	**x**
Net1	categ 1	0.002	33.433
	categ 2	0.9499	31.82165
	categ 3	0.0259	34.36765
	categ 4	0.0222	67.7437
Net2	categ 1	0.119	29.5135
	categ 2	0.8434	28.2539
	categ 3	0.0355	34.68925
	categ 4	0.002	67.067
Net3	categ 1	0.8319	5.6313
	categ 2	0.1665	5.5777
	categ 3	0.000	33.5
	categ 4	0.0016	67.05

From where the fuzzy input sets are obtained, as fuzzy antecedent:

20cm = {0.0/33.43;0.95/31.82;0.02/34.36; 0.02/67.74}
40cm = {0.12/29.51; 0.84/28.25; 0.03/34.68; 0.0/67.06}
60cm = {0.83/5.63; 0.16/5.57; 0.0/33.5; 0.0/67.05}

They will operate with the relational fuzzy sets, which in turn have the same structure as the consequent diffuse, corresponding to the output, according to Figure. 4.

Nippers = {1/0; 0.03/35; 0.0/67; 0/100}
Screwdriver = {0/0; 1/35; 0.01/67; 0/100}
Scalpel = {0/0; 0.0/35; 1/67; 0.01/100}
Scissors = {0/0; 0/35; 0.01/67; 1/100}

For the case, the computation of the Mamdani algorithm for the proposed example is illustrated. The total extent of the calculations involves finding the 12 βn and Bn, for each possible combination of category detection and/or confusion of them.

Step 1.
β1 = max({0.0; 0.95; 0.02; 0.02}^{1; 0.03; 0.0; 0.0}) = 0.03
β2 = max({0.0; 0.95; 0.02; 0.02}^{0; 1; 0.01; 0}) = 0.95
β3 = max({0.0; 0.95; 0.02; 0.02}^{0; 0.0; 1; 0.01}) = 0.02
β4 = max({0.0; 0.95; 0.02; 0.02}^{0; 0; 0.01; 1}) = 0.01
β5 = max({0.12; 0.84; 0.03; 0.0}^{1; 0.03; 0.0; 0.0}) = 0.12
β6 = max({0.12; 0.84; 0.03; 0.0}^{0; 1; 0.01; 0}) = 0.84
β7 = max({0.12; 0.84; 0.03; 0.0}^{0; 0.0; 1; 0.01}) = 0.03
β8 = max({0.12; 0.84; 0.03; 0.0}^{0; 0; 0.01; 1}) = 0.01
β9 = max({0.83; 0.16; 0.0; 0.0}^{1; 0.03; 0.0; 0.0}) = 0.83
β10 = max({0.83; 0.16; 0.0; 0.0}^{0; 1; 0.01; 0}) = 0.16
β11 = max({0.83; 0.16; 0.0; 0.0}^{0; 0.0; 1; 0.01}) = 0.0
β12 = max({0.83; 0.16; 0.0; 0.0}^{0; 0; 0.01; 1}) = 0.0

Step 2.
B1 = 0.03^1; 0.03; 0.0; 0} ={0.03; 0.03; 0.0; 0}
B2 = 0.95^0; 1; 0.01; 0} ={0; 0.95;0.01;0}
B3 = 0.02^0; 0.0; 1; 0.01} ={0; 0.0; 0.02; 0.01}
B4 = 0.01^0; 0; 0.01; 1} ={0; 0; 0.01; 0.01}
B5 = 0.12^1; 0.03; 0.0; 0} ={0.12; 0.03; 0.0; 0}
B6 = 0.84^{0; 1; 0.01; 0} ={0; 0.84;0.01;0}
B7 = 0.03^{0; 0.0; 1; 0.01} ={0; 0.0; 0.03; 0.01}
B8 = 0.01^{0; 0; 0.01; 1} ={0; 0; 0.01; 0.01}
B9 = 0.83^{1; 0.03; 0.0; 0} ={0.83; 0.03; 0.0; 0}
B10 = 0.16^{0; 1; 0.01; 0} ={0; 0.16;0.01;0}
B11 = 0.0^{0; 0.0; 1; 0.01} ={0; 0.0; 0.0; 0.0}
B12 = 0.0^{0; 0; 0.01; 1} ={0; 0; 0.0; 0.0}

Step 3.
B’ ={0.83; 0.95;0.03;0.01}

βdepth = 0.5
Bdepth = 0.5^{1; 1; 0.01; 0} ={0.5; 0.5; 0.01; 0.0}

It can be seen that because the depth value crosses by 0.5 when intersecting two functions of fairness in the boundary does not alter the result of B’. Therefore, it is to this result (B’) that the method of defuzzification by center of gravity is applied, by eq. (1).

$$y' = \frac{\sum_{j=1}^{F} \mu_{B'}(y_j) \cdot y_j}{\sum_{j=1}^{F} \mu_{B'}(y_j)} = 21.52 \quad (1)$$

This result corresponds to a membership value of 0.674 in the screwdriver category and 0.326 in the nippers category, where it is possible to decrease the influence of the third network by 50%. As a complementary example, in the case where the confusion of the third network went to the scissors category with a B’ ={0,02;

0,95;0,03;0,71}, he computation with Bdepth={0,5; 0,5; 0,01; 0,0} would generate a center of gravity at 50.6 corresponding to a membership of 0.485 in a screwdriver and 0 in scissors, with a reduction of about 70% in the influence of the prediction of the third network. This case is hypothetical because the class confusion analysis allowed to establish the relations of the membership functions, where there are no confusions between these two classes, so that their intersection is almost null (see Figure. 4). The result of the predicted design of the hybrid neuro-convolutional architecture, based on depth. It is appreciated that the images used are recognized satisfactorily when the RGB-D camera is approached towards the objects of interest. For the case, the overall error in the detection by variation of distances is 9.81%.
Where for the last example of the nippers of Figure. 6 (lower right), there is the capture ratio shown in Table 4. It is observed that the final fuzzy weighting improves the overall performance of the network.

Table 4. Result by Depth Identification.

Depth	Prediction Category
CCN1	0.93
CCN2	0.861
CCN3	0.43
DEPTH	0.87
FUZZY	0.902

The results obtained are demarcated by the range of coverage projected for a didactic robotic arm, whose dimensions allowed to set the depth ranges, in conjunction with the RGB-D capture camera used.

4 Conclusions

The proposed novel hybrid network architecture presents a solution to the problem of identifying objects through dynamic system that vary the image capture distance. The weighting achieved allows the clear recognition of training objects in the range of such training, improving the overall performance of the network, even above that of a trained network with an expanded database with different depths. For the latter case, the training is difficult for the network due to the variation of parameters derived from the depth changes, failing to obtain more than 76.2% accuracy with networks of 4 layers of convolution, evidencing the advantage of the hybrid network.

The final architecture developed presents an overall functionality of a conventional convolutional neural network, however, adding an additional layer of diffuse output enhances its object recognition capability in mobile robotic applications. Applications of this development can be used in tracking algorithms of objects subject to displacement, in comparison to the one presented where the camera is approaching or moving away from it.

References

[1] J. Schmidhuber, Deep learning in neural networks: An overview, *Neural Networks*, **61** (2015), 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

[2] Z. Fadlullah, F. Tang, B. Mao, N. Kato, O. Akashi, T. Inoue, K. Mizutani, State-of-the-Art Deep Learning: Evolving Machine Intelligence Toward Tomorrow's Intelligent Network Traffic Control Systems, *IEEE Communications Surveys & Tutorials*, **19** (2017), no. 4, 2432-2455. https://doi.org/10.1109/comst.2017.2707140

[3] Chung-Hsun Sun, Ying-Jen Chen, Yin-Tien Wang, Sheng-Kai Huang, Sequentially switched fuzzy-model-based control for wheeled mobile robot with visual odometry, *Applied Mathematical Modelling*, **47** (2017), 765-776. https://doi.org/10.1016/j.apm.2016.11.001

[4] Mohamed Slim Masmoudi, Najla Krichen, Mohamed Masmoudi, Nabil Derbel, Fuzzy logic controllers design for omnidirectional mobile robot navigation, *Applied Soft Computing*, **49** (2016), Pages 901-919. https://doi.org/10.1016/j.asoc.2016.08.057

[5] Fatemeh Fathinezhad, Vali Derhami, Mehdi Rezaeian, Supervised fuzzy reinforcement learning for robot navigation, *Applied Soft Computing*, **40** (2016) 33-41. https://doi.org/10.1016/j.asoc.2015.11.030

[6] Barmak Baigzadehnoe, Zahra Rahmani, Alireza Khosravi, Behrooz Rezaie, On position/force tracking control problem of cooperative robot manipulators using adaptive fuzzy backstepping approach, *ISA Transactions*, **70** (2017), 432-446. https://doi.org/10.1016/j.isatra.2017.07.029

[7] A. Krizhevsky, I. Sutskever and G. E. Hinton, Imagenet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems*, 2012, 1097-1105.

[8] M. Mancini, S. R. Bulò, E. Ricci and B. Caputo, Learning Deep NBNN Representations for Robust Place Categorization, *IEEE Robotics and Automation Letters*, **2** (2017), no. 3, 1794-1801. https://doi.org/10.1109/lra.2017.2705282

[9] J. Redmon and A. Angelova, Real-time grasp detection using convolutional neural networks, *2015 IEEE International Conference on Robotics and Automation (ICRA)*, (2015), 1316-1322. https://doi.org/10.1109/icra.2015.7139361

[10] Zhichao Wang, Zhiqi Li, Bin Wang, Hong Liu, Robot grasp detection using multimodal deep convolutional neural networks, *Advances in Mechanical Engineering*, **8** (2016), no. 9. https://doi.org/10.1177/1687814016668077

[11] S. Praveena and S. P. Singh, Hybrid clusteing algorithm and Neural Network classifier for satellite image classification, *2015 International Conference on Industrial Instrumentation and Control (ICIC)*, (2015), 1378-1383. https://doi.org/10.1109/iic.2015.7150963

[12] P. R. Futane and R. V. Dharaskar, Video gestures identification and recognition using Fourier descriptor and general fuzzy minmax neural network for subset of Indian sign language, *2012 12th International Conference on Hybrid Intelligent Systems (HIS)*, (2012), 525-530. https://doi.org/10.1109/his.2012.6421389

[13] G. D. Wu and Po-Jen Wu, Type-2 fuzzy neural network for voice activity detection, *2016 International Conference on Fuzzy Theory and Its Applications (iFuzzy)*, (2016), 1-4. https://doi.org/10.1109/ifuzzy.2016.8004927

[14] D. Krleža and K. Fertalj, Graph Matching Using Hierarchical Fuzzy Graph Neural Networks, *IEEE Transactions on Fuzzy Systems*, **25** (2017), no. 4, 892-904. https://doi.org/10.1109/tfuzz.2016.2586962

[15] N. Baklouti and A. M. Alimi, Interval type-2 beta fuzzy neural network for wheeled mobile robots obstacles avoidance, *2017 International Conference on Advanced Systems and Electric Technologies (IC_ASET)*, (2017), 481-486. https://doi.org/10.1109/aset.2017.7983740

[16] Y. Guo, W. Wang and S. Wu, Research on robot path planning based on fuzzy neural network and particle swarm optimization, *2017 29th Chinese Control And Decision Conference (CCDC)*, (2017), 2146-2150. https://doi.org/10.1109/ccdc.2017.7978870

[17] Y. Deng, Z. Ren, Y. Kong, F. Bao and Q. Dai, A Hierarchical Fused Fuzzy Deep Neural Network for Data Classification, *IEEE Transactions on Fuzzy Systems*, **25** (2017), no. 4, 1006-1012. https://doi.org/10.1109/tfuzz.2016.2574915

[18] E. P. Ijjina and C. K. Mohan, Human action recognition based on motion capture information using fuzzy convolution neural networks, *2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR)*, (2015), 1-6. https://doi.org/10.1109/icapr.2015.7050706

[19] Ho-Joon Kim, J. S. Lee and J. H. Park, Dynamic hand gesture recognition using a CNN model with 3D receptive fields, *2008 International Conference on Neural Networks and Signal Processing*, (2008), 14-19. https://doi.org/10.1109/icnnsp.2008.4590300

[20] Zeiler M.D., Fergus R. (2014) Visualizing and Understanding Convolutional Networks, Chapter in *Computer Vision – ECCV 2014*, Vol. 8689, D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (eds.), Lecture Notes in Computer Science, Springer, Cham. https://doi.org/10.1007/978-3-319-10590-1_53

[21] Babuska Robert, Fuzzy and Neural Control DISC Course Lecture notes, Delft University of Technology, 2004.

Received: November 11, 2017; Published: December 15, 2017