International Journal of Electrical and Computer Engineering (IJECE)
Vol. 10, No. 1, February 2020, pp. 1017~1026
ISSN: 2088-8708, DOI: 10.11591/ijece.v10i1.pp1017-1026

ResSeg: Residual encoder-decoder convolutional neural network for food segmentation

Javier O. Pinzón-Arenas, Robinson Jiménez-Moreno, Cesar G. Pachón-Suescún
Faculty of Engineering, Nueva Granada Military University, Colombia

Article Info

Article history:

Received Jun 2, 2019
Revised Sep 2, 2019
Accepted Sep 27, 2019

Keywords:

Encoder-decoder CNN
Food recognition
Residual layers
SegNet
Semantic segmentation

ABSTRACT

This paper presents the implementation and evaluation of different convolutional neural network architectures focused on food segmentation. To perform this task, it is proposed the recognition of 6 categories, among which are the main food groups (protein, grains, fruit, and vegetables) and two additional groups, rice and drink or juice. In addition, to make the recognition more complex, it is decided to test the networks with food dishes already started, i.e. during different moments, from its serving to its finishing, in order to verify the capability to see when there is no more food on the plate. Finally, a comparison is made between the two best resulting networks, a SegNet with architecture VGG-16 and a network proposed in this work, called Residual Segmentation Convolutional Neural Network or ResSeg, with which accuracies greater than 90% and interception-over-union greater than 75% were obtained. This demonstrates the ability, not only of SegNet architectures for food segmentation, but the use of residual layers to improve the contour of the segmentation and segmentation of complex distribution or initiated of food dishes, opening the field of application of this type of networks to be implemented in feeding assistants or in automated restaurants, including also for dietary control for the amount of food consumed.

Corresponding Author:

Javier Orlando Pinzón Arenas,
Mechatronics Engineering Program, Faculty of Engineering,
Nueva Granada Military, University,
Carrera 11 #101-80, Bogotá D.C., Colombia.
Email: u3900231@unimilitar.edu.co

1. INTRODUCTION

The recognition of patterns applied to food is a topic that has begun to take importance within machine vision systems, mainly focused on calorie control [1] or diet control [2]. However, that application has been a challenge because the dishes do not contain characteristic shapes or the ingredients used can differ by drastically changing the visual characteristics of a type of food. For the execution of this task, several developments have been implemented, such as those presented in [3-5], but the main problem is that these systems depend on the dish containing only one type of food, recognizing a specific dish without discriminating the ingredients [6] or being in a controlled or known environment [7, 8].

To increase the robustness of the food recognition systems, other techniques have begun to be applied in a way that allows recognizing different types of food in a single dish. Some developed examples are presented in [9] and [10]. The first one makes use of pairwise statistics to recognize different categories of food on a plate by means of the exploitation of local features, but it depends on being in a controlled environment with a white background. In the second one, a combination of segmentation of objects with perceptual similarities is used, in such a way that the algorithm recognizes and segments food types with respect to their characteristics, achieving an accuracy of 44% in the segmentation of 32 meals. Although these techniques already discriminate parts of the same dish, the performance is very low, which is

why researchers began to include artificial intelligence in this area, more specifically Deep Learning (DL) techniques [11].

Within DL, there is a technique called Convolutional Neural Network (CNN) [12, 13], which has had an exponential evolution, not only in its performance in the recognition of patterns in images but in its variations and application. In the work described in [14], a comparison is made between region-based CNN and a Deep CNN, to segment and locate food in a photo, while in [15], a Deep CNN is used with a variation in the structure of the fully connected layers to segment the food along with a multi-scale CNN to estimate its depth, achieving results with ranges between 70% and 75% accuracy in the estimation of food quantity.

On the other hand, other architectures have been developed to achieve an improvement in the segmentation of objects, called Encoder-Decoder CNNs, or commonly known as SegNet [16], which is a CNN that consists of two stages. The first stage consists of an encoder in charge of generating the recognition of the object, however, it does not contain fully connected layers, i.e. its last convolution layer is connected directly to the second stage, which is a mirror of the encoder, called decoder, adding a direct connection between each section of the encoder's downsampling with its respective part of the decoder's upsampling, allowing having a better characterization of the image in the last layers. This type of network has had a great performance in tasks related to the segmentation of objects [17-19], even in the segmentation of medical images [20-23], which have the characteristic of not having sections with a specific shape or totally amorphous. However, this network has not been widely used in the task of food segmentation, for this reason, this work explores the possibility of being used and demonstrate its performance in this task. Although the food segmentation developments are mainly focused on the dietary control, this work expands the use of these systems to know the existence or not of food on a dish, so that it can be applied in future work to developments that require knowing the percentage of current food or autonomous systems of assisted feeding. Likewise, different architectures are implemented and evaluated to analyze their results with respect to an architecture proposed in the state of the art, exploring the use of residual layers in conjunction with the SegNet, which is called in this work ResSeg.

The work is divided into 4 sections, including the present introduction. In section 2, the database prepared along with the proposed architectures is presented. Section 3 describes the results obtained from the training and testing of the networks, taking into account the use and non-use of the background label. Finally, in section 4, the conclusions reached are given.

2. METHODOLOGY

The work done focuses on the segmentation of 6 food groups, for this case, from the lunch meal, within which are the 4 main groups of foods (Protein, Vegetables, Fruits, and Grains) plus two subgroups, where Juice and Rice are located. This is done since, in the case of rice, in the common food menu of the Colombian region it is found in most dishes, and for juice, because it is the liquid part of lunch. It should be noted that the soup, which is also an essential element in the food of the region, is not taken into account. For this, the construction of our own dataset is made, as well as the proposal of different architectures and their comparison with the VGG-16 for semantic segmentation. Next, the development of the work is exposed.

2.1. Database

The elaborated dataset consists of images of basic food dishes, or commonly called "Executives" in the region. This dish consists of a portion of rice, a type of grain, a protein, a portion of fruit, a glass of juice, and mostly a portion of salad, and are taken on backgrounds with simple and complex textures. Additionally, not only images of freshly prepared dishes are taken, but as the person consumes it, pictures of it are taken, increasing its complexity for recognition, because in the dish, there are residues of sauces and the food portions begin to be separated or mixed. The pictures are taken in two restaurants, without control of the lighting and without a fixed distance between the plate and the camera, but mostly taken around 55 cm of distance. These images are adjusted to a standard size of 480x360 pixels, to avoid a high computational cost in training if higher resolutions are used. Each photo is manually labeled, obtaining a total of 236 images, where 200 are used for training and 36 for validation of networks. An example of the dataset can be seen in Figure 1 along with its labeling. It should be noted that no data augmentation has been performed.

2.2. Proposed architectures

In order to evaluate the food segmentation capacity with a small database as the one elaborated here, different architectures of CNNs were proposed, using configurations such as SegNet or Encoder-Decoder with variation in depth, and variants of this architecture, which contain residual layers. This was done in order to observe their performance and quality of the segmentation of each category. Each of the architectures is shown in Figure 2 and are described below.

(a)

(b)

Figure 1. Examples of the dataset, where (a) a recently prepared dish and (b) already started to be eaten, with their respective labels

The first architecture is a SegNet of depth 3 (or SegNet D3), i.e. it is formed by an Encoder network that consists of 3 sets of layers. Each set contains 2 groups of convolution, in other words, a group is composed of a convolution layer, a batch normalization layer and a ReLU layer. At the end of each set, a downsampling or maxpooling layer is added. The Encoder is followed by a Decoder network that is basically a mirror of the encoder, with the difference that instead of having downsampling, the decoder has upsampling layers called unpooling layer. In order to improve the delineation of the segmentation by means of the retention of the details of the encoder's early layers, the indices of each pooling layer are interconnected with the indices of their respective mirror unpooling layer, as defined in [16]. The second architecture, called SegNet D4, has a similar configuration as the previous one, with a slight variation in its depth, being of 4, that is, with an additional set of layers in both the encoder and decoder. For architecture 3, which is a modification of the SegNet D3, the filters' size of the first set of the encoder and the last set of the decoder were varied, establishing them in 5x5 pixel squares, so that the network could learn in a better way textures and internal shapes of the food. Likewise, the number of filters were increased in some layers.

In order to explore architectures different from conventional ones, it was proposed to use residual layers within each layer set. As reported in [24], the addition of this type of layer improves the accuracy of the network in conjunction with the quality of the segmentation with respect to the contour of the object thanks to the transfer of the features of early layers to deeper layers. For this reason, the following architectures, called E-Residual v.1 and v.2, consist of a SegNet of depth three with sets of layers composed of 3 groups of convolution, adding residual layers in each set of the encoder. However, due to the interconnectivity of the encoder with the decoder, the residual layer input is taken from the first group of convolution of the set, and its output is connected to the pooling layer input. The connections can be observed more graphically in Figure 3, where the output of the connection is the ReLU function. In this figure, "conv" refers to a convolution layer and "B-norm" to a batch normalization layer. The main difference between these two networks is the number of filters used per layer.

To further strengthen the previous architectures, residual connections are added in the decoder. This variation is named Residual Segmentation Convolutional Neural Network or ResSeg, which has a depth of 3, or ResSeg D3. Finally, the last proposed architecture is a ResSeg with a depth of 5 sets (ResSeg D5). Its main characteristic that, in the first set of the encoder and last of the decoder, residual branches are not used, and filters of size 3 are used in the first convolution layer, except in stage 4, because it is necessary to adjust the output volume to avoid loss of information, due to the size of the image.

It should be noted that for the convolution layers with 3x3 filter size in every architecture, padding of 1 is used, in such a way that the size of the input volume of the layer is maintained. The network with which it will be compared to the performance of all proposed architectures is a SegNet version of the VGG-16 [25], since its depth, and in general, its architecture is similar to the proposed ones, but with a greater number of convolution layers and learning filters.

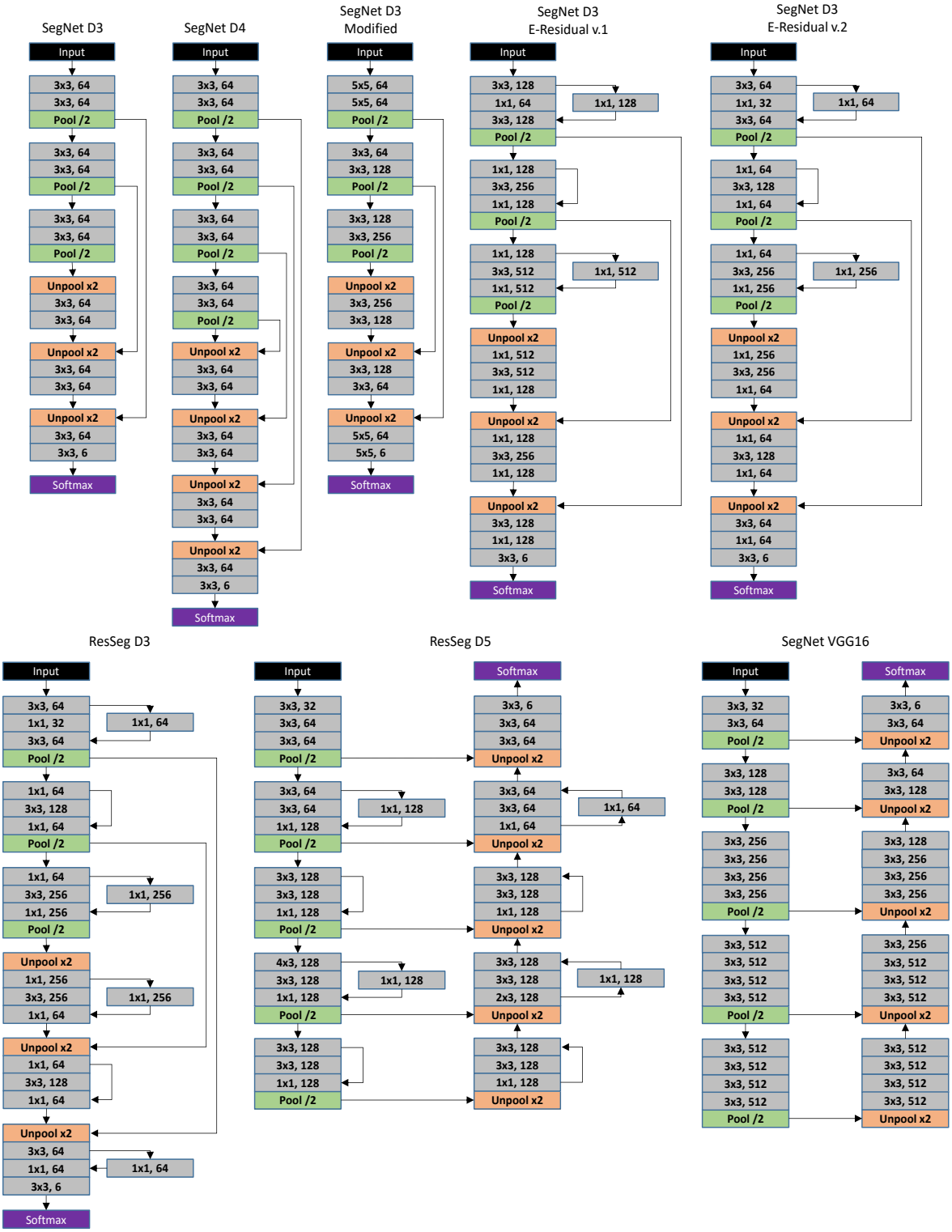


Figure 2. Proposed architectures

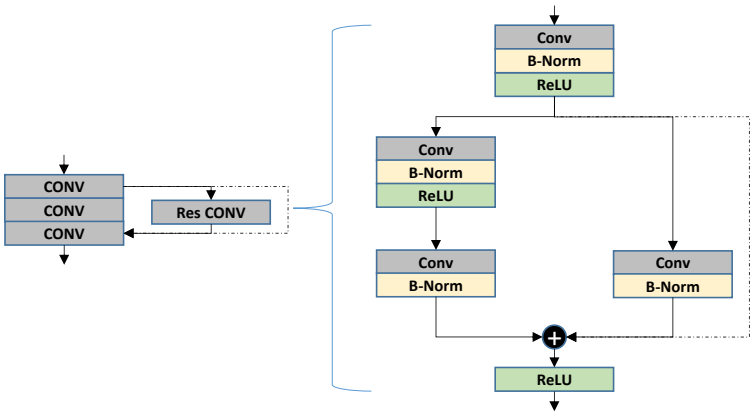


Figure 3. Residual connection structure

3. RESULTS AND DISCUSSIONS

3.1. Performance without labeled background

For the purpose of comparing the performance of the architectures, identical training parameters are set, so that they can learn under the same conditions. Taking into account the above, the initial learning rate is set to 10^{-4} with a drop factor of 0.5 per 100 epochs, using a batch size of 2 and a total of 400 epochs of training. Those parameters were set by doing iterative tests, with which better results were obtained during the training. With this, the behaviors shown in Figure 4 are obtained, where no network surpasses an accuracy of 50%, even having the VGG16 as the worst performer, with 30% accuracy.

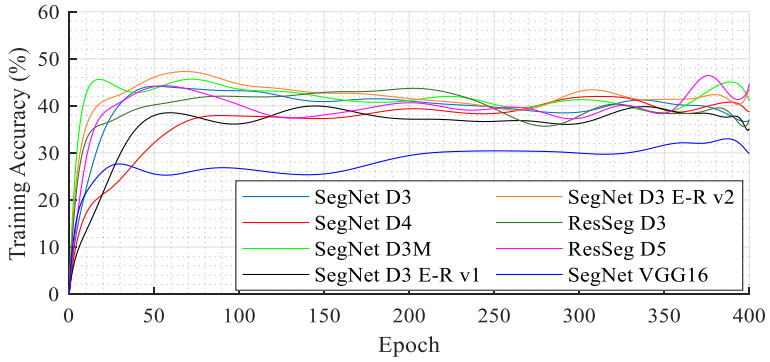


Figure 4. Training behavior of each architecture, using a database without labeled background

Due to the low performance of the networks during their training, it is necessary to observe what they learned to understand their behaviors. For this, tests are made with test images, obtaining results like the one in Figure 5, where, despite being able to segment each type of food, they are not able to eliminate the background or parts of the dishes, causing large amounts of false positives to be generated from all foods, which makes the network inefficient.

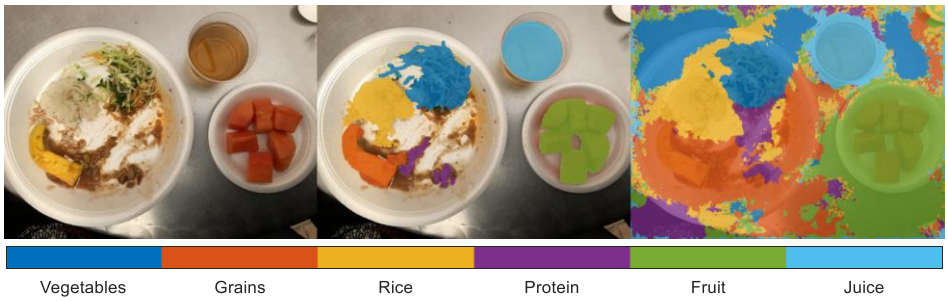


Figure 5. Comparison between ground truth and segmentation obtained from ResSeg D3

3.2. Performance with labeled background

To solve the above problem, it is proposed to add an additional category called "Background", where each unlabeled part of the image becomes part of this category, generating images like the example shown in Figure 6. With this modification, it is proceeded to perform the training of each network again. However, a slight modification is made in the learning rate parameter, using an initial value of 10^{-3}, with a drop ratio of 0.5 per 150 epochs, in such a way that the networks start with a rapid learning process, and then fine-tuning the parameters learned every certain number of epochs. This modification was decided after looking at the initial behavior of the networks compared to the first learning rate used, seeing that with a smaller learning rate, the networks tended to obtain accuracies lower than 70% after 200 epochs. With these parameters, a training behavior like the one shown in Figure 7 is obtained, having as the two best networks the ResSeg of depth 5 and the SegNet with VGG-16 architecture, with 94.1% and 95.5% accuracy, respectively, surpassing by more than 3% the modified SegNet D3, which was the only one to achieve more than 90% among the remaining networks.

Figure 6. Ground truth with the additional category “Background”

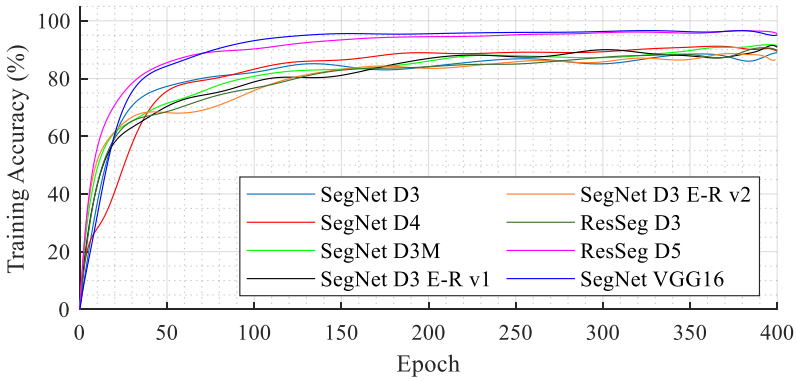


Figure 7. Training behavior of each architecture, using a database with labeled background

3.3. Test performance of final trained architectures

Since the architectures had better behavior during their training, it is proceeded to verify their performance with the test images. Table 1 shows the results obtained with each architecture. The architectures with the lowest performance were the SegNets to which the residual layers were added only in the encoder stage. On the other hand, there are networks with percentages of accuracy greater than 90%, with SegNet D4, modified D3, and ResSeg D5, with slight differences. However, an important factor for its evaluation is presented in the intercept over union (IoU), where the ResSeg obtained a higher relation, i.e. how well the network classifies the pixels of the classes, given by (1), where TP are the true positives, FP the false positives and the FN the false negatives, being the value shown in the table the average of all the classes of the evaluated dataset.

Table 1. Architectures performance with the test database

Network	SegNet D3	SegNet D4	SegNet D3 M	SegNet D3 E-R v.1	SegNet D3 E-R v.2	ResSeg D3	ResSeg D5	SegNet VGG16
Mean Accuracy (%)	89.61	90.06	90.35	87.85	86.71	88.17	90.43	94.86
Mean IoU	0.635	0.663	0.661	0.638	0.631	0.642	0.756	0.821
Mean BF Score	0.547	0.568	0.590	0.567	0.554	0.567	0.707	0.768
Mean Processing Time (ms)	163.9	183.3	260.3	317.6	203.3	210.0	264.1	354.1

$$IoU = \frac{TP}{TP+FP+FN} \quad (1)$$

Making a comparison between the two proposed ResSeg architectures, it can be seen that, with the increase in the depth of the network, the average accuracy increased by 2.26%. Likewise, the mIoU obtained a significant improvement, increasing by 0.114, like the parameter BF (boundary F1) which is defined as the precision in the alignment of the predicted boundary with respect to the ground truth, growing by 0.14, observing an increase in the processing time of only 50 ms. With this, it can be inferred that by increasing the depth to a certain degree, it can be easily reached the performance of the SegNet VGG16, even being able to have a better processing time.

A comparison of the architectures in different dishes and environments can be seen in Figure 8, where there are freshly served dishes, started to be eaten and finished. Similarly, tables with flat color and complex textures are used in such a way that it is more difficult to differentiate it from food. When the dishes are freshly served, such as number 4, the networks tend to generate a good classification and segmentation of the food, although they have parts of the table that are recognized as some type of food, except ResSeg D5 and VGG16. In the same way, when the plate is almost empty (column 5), they generate a good classification, even of parts not labeled in the ground truth, although they also manage to label leftovers of sauce located in the upper part of the plate. In the first column, the networks tend to be wrong mainly in the segmentation of the juice, mainly by its color, with the exception of the last two architectures, even though part of the meat sauce was labeled as meat in the ground truth, these two were able to identify which was the protein and separate it from the sauce.

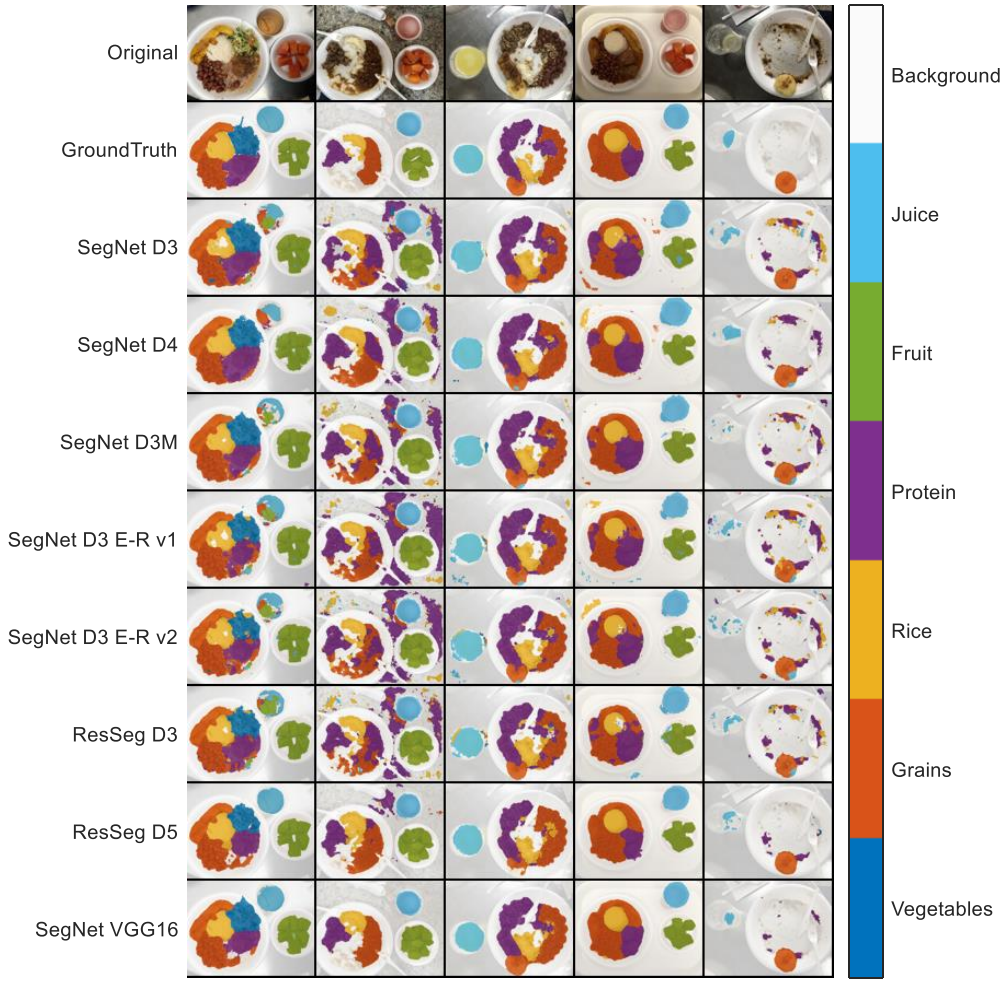


Figure 8. Different tests of the architectures with the test dataset

The most complex image is located in column 2, where there is a plate that has already been started and a background with a complex texture. In this, the first 6 architectures generate a large number of false positives, especially protein due to the dark color of the texture. The ResSeg D5 network, although its amount of false positives is largely less than other networks, tends to be wrong where the candy is (which is not part of any category). On the other hand, VGG16 was able to discriminate this and avoid in large quantity false positives caused by the environment. In general terms, the two best architectures are the ResSeg D5 and SegNet VGG16, which are able to eliminate the noise that does not belong to the dish to a large extent and to segment with good precision the types of food.

On the other hand, although the VGG16 has better overall performance, when making the comparison between this network and the ResSeg D5, especially with freshly served dishes, the ResSeg has a better behavior regarding the delineation of the contours of the segmented sections, as shown in Figure 9. The ResSeg is able to segment empty spaces between parts of the same food, as can be seen in the left part of the dish, where a noodle leaves an empty space, which is filled by the VGG16. It even has a better segmentation of the vegetable category. This allows you to have a better idea of the amount of food that is actually on the plate.

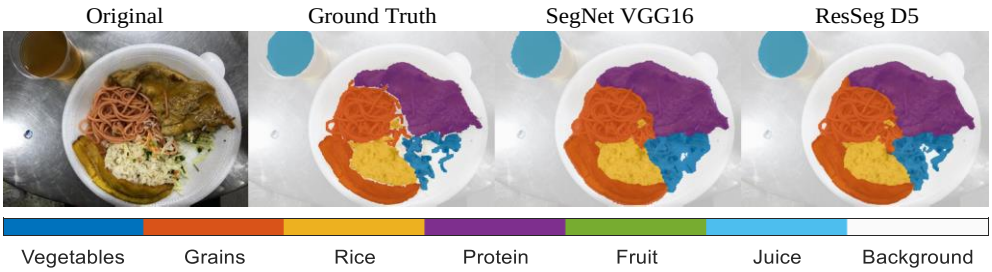
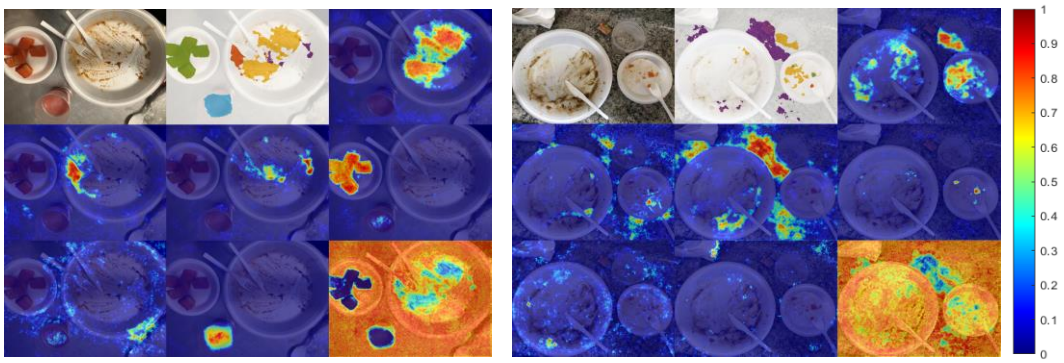


Figure 9. Comparison of boundary segmentation performance between ResSeg D5 and VGG16

3.4. Cases of wrong segmentation for ResSeg D5

The ResSeg D5, although it has a good behavior especially when there is food still on the plate, has a difficulty when there is no food but there are residues of sauces that can simulate the texture of the rice by the division and color of the dish, causing it to generate a confusion of quantity of food. It can be seen in Figure 10a, where according to their activations, the network activates areas mainly where the sauce is located. This can be caused by convolutions from pixel to pixel (filters of size 1), since they can divert their learning towards active but little relevant parts, taking into account that prior to these there are only 2 layers of convolution and do not use a large number of filters in the layers, considering the number of possible details that may be in the empty plates.

Another similar situation happens but this time when complex backgrounds are presented, such as the one presented in Figure 10b. Here, the network, despite having been able to correctly identify the main food dish with some mistakes in the fruit plate, presents segmentation of the protein category in the table, due to the texture it has, making it look like ground meat, also being activated where the sweet is, and inside the glass as if it were rice. Although this does not happen to a large extent, when lighting is low, this error tends to appear.



(a) (b)

Figure 10. Cases of wrong segmentation

4. CONCLUSIONS

In this paper, it was explored the use of residual layers for the task of food segmentation in convolutional neural networks, within which an architecture with performance close to one of the most used networks in segmentation was implemented, which was named ResSeg. Different architectures with different configurations were structured in the location of the residual layers, making comparisons with basic segmentation networks, showing that having a shallow depth, the residual layers make the architecture perform less than those architectures that do not use them, even with the same depth, as shown in Table 1. On the other hand, when increasing the depth of the network, in the case of comparison between ResSeg D3 and D5, the performance improves greatly, by more than 2% in the average accuracy, and up to 11% the performance of the mIoU.

The correct labeling of the database plays a fundamental role in the training of architectures, since, as evidenced in section 3.1, not having a label corresponding to the background, makes the performances of these be critically impaired, taking into account everything not labeled as parts of the other categories. For this reason, it is important to include an additional label that includes the parts not related to the categories in the image. Although the ResSeg D5 has a high performance, exceeding 90% accuracy, it still remains below the SegNet VGG-16, however, when processing images with freshly served food, the contour of the segmented sections improves, avoiding segmenting empty spaces between the same types of food and better delineating the contours between each category, as shown in Figure 9. In future work, it is proposed to increase the capacity of the network by adding a greater depth in each phase of the encoder/decoder with 3x3 filters. This can improve its accuracy, since it would allow it to learn in a better way the characteristics of the complex textures of the environment to avoid the confusion that may be generated by the layers with 1x1 filters, implementing new configurations.

ACKNOWLEDGEMENTS

The authors are grateful to the Nueva Granada Military University, which, through its Vice-chancellor for research, finances the present project with code IMP-ING-2935 (being in force 2019-2020) and titled “Flexible robotic prototype for feeding assistance”, from which the present work is derived.

REFERENCES

[1] P. Pouladzadeh, G. Villalobos, R. Almaghrabi and S. Shirmohammadi, "A novel SVM based food recognition method for calorie measurement applications," in *2012 IEEE International Conference on Multimedia and Expo Workshops, IEEE*, pp. 495-498, 2012.

[2] V. Bruno and C. J. Silva Resende, "A survey on automated food monitoring and dietary management systems," *Journal of health & medical informatics*, vol. 8(3), 2017.

[3] M. Bosch, F. Zhu, N. Khanna, C. J. Boushey and E. J. Delp, "Combining global and local features for food identification in dietary assessment," in *Image Processing (ICIP), 2011 18th IEEE International Conference on, IEEE*, pp. 1789-1792, 2011.

[4] Y. Matsuda, H. Hoashi and K. Yanai, "Recognition of multiple-food images by detecting candidate regions," in *Multimedia and Expo (ICME), 2012 IEEE International Conference on, IEEE*, pp. 25-30, 2012.

[5] F. Kong and J. Tan, "DietCam: Automatic dietary assessment with mobile camera phones," *Pervasive and Mobile Computing*, vol. 8(1), pp. 147-163, 2012.

[6] L. Bossard, M. Guillaumin and L. Van Gool, "Food-101–mining discriminative components with random forests," in *European Conference on Computer Vision, Springer, Cham*, pp. 446-461, 2014.

[7] W. Zhang, Q. Yu, B. Siddiquie, A. Divakaran and H. Sawhney, ""Snap-n-Eat" food recognition and nutrition estimation on a smartphone," *Journal of diabetes science and technology*, vol. 9(3), pp. 525-533, 2015.

[8] M. Y. Chen, *et al.*, "Automatic Chinese food identification and quantity estimation," in *SIGGRAPH Asia 2012 Technical Briefs*, ACM, pp. 29, 2012.

[9] S. Yang, M. Chen, D. Pomerleau and Sukthankar R., "Food recognition using statistics of pairwise local features," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE*, pp. 2249-2256, 2010.

[10] F. Zhu, M. Bosch, N. Khanna, C. J. Boushey and E. J. Delp, "Multilevel segmentation for food classification in dietary assessment," in *Image and Signal Processing and Analysis (ISPA), 2011 7th International Symposium on, IEEE*, pp. 337-342, 2011.

[11] Y. LeCun, Y. Bengio and G. Hinton, "Deep learning," *Nature*, vol. 521(7553), 436, 2015.

[12] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European conference on computer vision, Springer, Cham*, pp. 818-833, 2014.

[13] S. Mallat, "Understanding deep convolutional networks," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374(2065), pp. 20150203, 2016.

[14] W. Shimoda and K. Yanai, "CNN-based food image segmentation without pixel-wise annotation," in *International Conference on Image Analysis and Processing, Springer, Cham*, pp. 449-457, 2015.

[15] A. Meyers, *et al.*, "Im2Calories: Towards an automated mobile vision food diary," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1233-1241, 2015.

[16] V. Badrinarayanan, A. Kendall, R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39(12), pp. 2481-2495, 2017.
[17] M. Kampffmeyer, A. B. Salberg and R. Jenssen, "Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 1-9, 2016.
[18] J. Cheng, Y. H. Tsai, W. C. Hung, S. Wang and M. H. Yang, "Fast and accurate online video object segmentation via tracking parts," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7415-7424, 2018.
[19] A. Kendall, V. Badrinarayanan and R. Cipolla, "Bayesian segnet: Model uncertainty in deep convolutional encoder-decoder architectures for scene understanding," *arXiv preprint arXiv:1511.02680*, 2015.
[20] J. Tang, J. Li and X. Xu, "Segnet-based gland segmentation from colon cancer histology images," in *2018 33rd Youth Academic Annual Conference of Chinese Association of Automation (YAC), IEEE*, pp. 1078-1082, 2018.
[21] P. Kumar, P. Nagar, C. Arora and A. Gupta, "U-Segnet: Fully convolutional neural network based automated brain tissue segmentation tool," in *2018 25th IEEE International Conference on Image Processing (ICIP), IEEE*, pp. 3503-3507, 2018.
[22] N. Ing, *et al*., "Semantic segmentation for prostate cancer grading by convolutional neural networks," in *Medical Imaging 2018: Digital Pathology*, International Society for Optics and Photonics, pp. 105811B, 2018.
[23] S. Alqazzaz, X. Sun, X. Yang and L. Nokes, "Automated brain tumor segmentation on multi-modal MR image using SegNet," *Computational Visual Media*, vol. 5(2), pp. 209-219, 2019.
[24] T. Pohlen, A. Hermans, M. Mathias and B. Leibe, "Full-resolution residual networks for semantic segmentation in street scenes," in *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on, IEEE*, pp. 3309-3318, 2017.
[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

BIOGRAPHIES OF AUTHORS

Javier Orlando Pinzón Arenas was born in Socorro-Santander, Colombia, in 1990. He received his degree in Mechatronics Engineering (Cum Laude) in 2013, Specialization in Engineering Project Management in 2016, and M.Sc. in Mechatronics Engineering in 2019, at the Nueva Granada Military University - UMNG. He has experience in the areas of automation, electronic control, and machine learning. Currently, he is studying a Ph.D. in Applied Sciences and working as a Graduate Assistant at the UMNG with emphasis on Robotics and Machine Learning.
E-mail: u3900231@unimilitar.edu.co

Robinson Jiménez Moreno was born in Bogotá, Colombia, in 1978. He received the Engineer degree in Electronics at the Francisco José de Caldas District University - UD - in 2002. M.Sc. in Industrial Automation from the Universidad Nacional de Colombia - 2012 and Ph.D. in Engineering at the Francisco José de Caldas District University - 2018. He is currently working as a Professor in the Mechatronics Engineering Program at the Nueva Granada Military University - UMNG. He has experience in the areas of Instrumentation and Electronic Control, acting mainly in Robotics, control, pattern recognition, and image processing.
E-mail: robinson.jimenez@unimilitar.edu.co

César Giovany Pachón Suescún was born in Bogotá, Colombia, in 1996. He received his degree in Mechatronics Engineering from the Pilot University of Colombia in 2018. Currently, he is studying his Master's degree in Mechatronics Engineering and working as Research Assistant at the Nueva Granada Military University with an emphasis on Robotics and Machine Learning.
E-mail: u3900259@unimilitar.edu.co