Ship detection using different segmentation methods

Ákos Gángoly, Gergely Kovács, Zoltán Gyenes

***Abstract*—Nowadays deep learning is one of the most important research topic in the science life. Image segmentation plays a big role in the research topic because of the big number of the applications that use this method. There are different types of segmentation techniques. Our goal was to implement and compare different types of segmentations solving a ship detection problem. We used:**

***Index Terms*—deep learning, image segmentation, transfer learning, segnet**

I. Introduction

In the field of computer vision semantic segmentation is one of the key problem and it is applied both at images and videos. The most important goal is always to understand the complete scene. Because of the growing number of the computer vision techniques there was an increase of the applications that use the imagery knowledge from the neighborhood. There are several applications to this topic, e.g.: autonomous driving [1], [2], [3] or image search engines [4].

Because of the deep learning revolution the old methods are not useable any more. At the segmentation problems usually the Convolutional Neural Networks (CNNs) are widely used [5], [6], [7], [8], because of the efficiency. In the past there were some articles to summarize the semantic segmentation techniques [9], [10] but the most expressive and most extensive work was considered the most recent datasets and provided details about deep learning techniques [11].

Our project aims to create a neural network to detect ships on satellite image. This task is important to track traffic and increase security against pirates.

There are many solutions for object detection on image but they were not yet implemented on satellite images to detect ships. Previous solutions used classic computer vision techniques. Instead of this we implemented segmentation techniques to solve this problem.

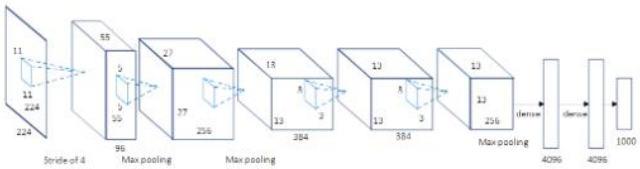
The rest of this paper will organized in the following order: Section II introduces the recently common used segmentation techniques. After that Section III. presents the Dataset that was used to the task. Section IV In Section V. the conclusion of the work will be presented and the opportunities of the further development will be defined.

II. Related Work

As it was mentioned, there are a lot of opportunities that were defined in the past to use the segmentation techniques [11]. In this section there will be overview about the AlexNet, VGG-16, GoogLeNet and ResNet methods.

A. AlexNet

The architecture of alexNet was defined by Krizhevsky [12]. It consists of five convolutional layers: max-pooling ones, Rectified Linear Units (ReLUs) as non-linearities, three fully-connected layers and dropout as you can see in II-1. Figure.



II-1. Figure: Architechture of AlexNet [12]

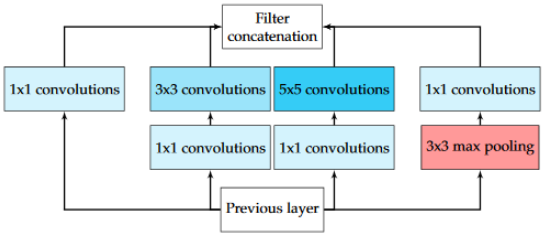
The AlexNet won the ILSVRC-2012 with on accuracy of 84.6 % against 73.8 %.

B. VGG

A group called VGG (Visual Geometry Group) from the Oxford university introduced a new CNN model called Visual Geometry Group (VGG). [13]. They submitted a model for ILSVRC-2013. This model had 16 weight layers (VGG-16) and it became very popular because of the achievement of 92.7 %. The main cause of the success were the stack of convolutional layers instead of few layers with big receptive fields.

C. GoogLeNet

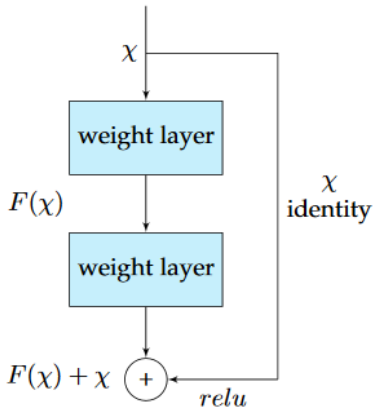
GoogLeNet nework won the LSVRC-2014. It was introduced by Szegedy [14]. This CNN architecture is characterized by its complexity, emphasized by the fact that it is composed by 22 layers and a newly introduced building block called inception module, it is shown in II-2. Figure.



II-2. Figure: Architecture of GooLeNet

D. ResNet

Microsoft was introduced a new Network called ResNet [15], and this netwok won the LSVRC-2016 with 96.4 % accuracy. The main difference of this network compared with the previous one was the depth of the network (152 layers) and the introduction of residual blocks (see in II-3. Figure). “The intuitive idea behind this approach is that it ensures that the next layer learns something new and different from what the input has already encoded (since it is provided with both the output of the previous layer and its unchanged input).”[11]Architecture of ResNet



II-3. Figure: Architecture of ResNet

III. DATASETS, DATA PREPROCESSING

IV. METHODS AND SOLUTIONS

V. CONCLUSION, FURTHER WORK

REFERENCES

[1] A. Ess, T. M̈uller, H. Grabner, and L. J. Van Gool, “Segmentation-based urban traffic scene understanding.” in BMVC , vol. 1, 2009,p. 2.

[2] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the kitti vision benchmark suite,” in 2012 IEEE Conference on Computer Vision and Pattern Recognition , June 2012, pp. 3354–3361

[3] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, “The cityscapes dataset for semantic urban scene understanding,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 3213–3223.

[4] J. Wan, D. Wang, S. C. H. Hoi, P. Wu, J. Zhu, Y. Zhang, and J. Li,“Deep learning for content-based image retrieval: A comprehen-sive study,” in Proceedings of the 22nd ACM international conference on Multimedia . ACM, 2014, pp. 157–166.

[5] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber,“Deep neural networks segment neuronal membranes in electron microscopy images,” in Advances in neural information processing systems , 2012, pp. 2843–2851.

[6] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, “Learning hierarchical features for scene labeling,” IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. 8, pp. 1915–1929, 2013.

[7] B. Hariharan, P. Arbel aez, R. Girshick, and J. Malik, “Simulta-neous detection and segmentation,” inEuropean Conference on Computer Vision . Springer, 2014, pp. 297–312.

[8] Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla: SegNet: A Depp Convolutional Encoder-Decoder Architecture for Image Segmentation, arXiv:1511.00561v3 [cs.CV] 10 Oct 2016

[9] H. Zhu, F. Meng, J. Cai, and S. Lu, “Beyond pixels:A comprehensive survey from bottom-up to semantic image segmentation and cosegmentation,”Journal of Visual Communication and Image Representation , vol. 34, pp. 12 – 27, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1047320315002035

[10] M. Thoma, “A survey of semantic segmentation,”CoRR , vol.abs/1602.06541, 2016. [Online]. Available: http://arxiv.org/abs/1602.06541

[11] A. Garcia-Garcia, S. Orts-Escolano, S.O. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez: A Rewiew on Deep Learning Techniques Applied to Semantic Segmentations. arXiv:1704.06857v1 [cs.CV] 22 Apr 2017, online Available: https://arxiv.org/pdf/1704.06857.pdf

[12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in neural information processing systems, 2012, pp. 1097–1105

[13] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv preprint arXiv:1409.1556 , 2014.

[14] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov,

[15] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.